A HISTORY OF WESTERN PHILOSOPHY

IN TWELVE ARTICLES

FROM THE

STANFORD ENCYCLOPEDIA OF PHILOSOPHY

(Excluding Bibliography and Notes)

Presocratic Philosophy

First published Sat Mar 10, 2007; substantive revision Mon Apr 4, 2016

The Presocratics were 6th and 5th century BCE Greek thinkers who introduced a new way of inquiring into the world and the place of human beings in it. They were recognized in antiquity as the first philosophers and scientists of the Western tradition. This article is a general introduction to the most important Presocratic philosophers and the main themes of Presocratic thought. More detailed discussions can be found by consulting the articles on these philosophers (and related topics) in the SEP (listed below). The standard collection of texts for the Presocratics is that by H. Diels revised by W. Kranz (abbreviated as DK). In it, each thinker is assigned an identifying chapter number (e.g., Heraclitus is 22, Anaxagoras 59); then the reports from ancient authors about that thinker's life and thought are collected in a section of "testimonies" (A) and numbered in order, while the passages the editors take to be direct quotations are collected and numbered in a section of "fragments" (B). Alleged imitations in later authors are sometimes added in a section labeled C. Thus, each piece of text can be uniquely identified: DK 59B12.3 identifies line 3 of Anaxagoras fragment 12; DK 22A1 identifies testimonium 1 on Heraclitus.

- 1. Who Were the Presocratic Philosophers?
- 2. The Milesians
- 3. Xenophanes of Colophon and Heraclitus of Ephesus
- 4. Parmenides of Elea
- 5. The Pythagorean Tradition
- 6. Other Eleatics: Zeno and Melissus
- 7. The Pluralists: Anaxagoras of Clazomenae and Empedocles of Acragas
- 8. Presocratic Atomism

1. Who Were the Presocratic Philosophers?

Our understanding of the Presocratics is complicated by the incomplete nature of our evidence. Most of them wrote at least one "book" (short pieces of prose writing, or, in some cases, poems), but no complete work survives. Instead, we are dependent on later philosophers, historians, and compilers of collections of ancient wisdom for disconnected quotations (fragments) and reports about their views (*testimonia*). In some cases, these sources had direct access to the works of the Presocratics, but in many others, the line is indirect and often depends on the work of Hippias, Aristotle, Theophrastus, Simplicius, and other ancient philosophers who did have such access. The sources for the fragments and testimonia made selective use of the material available to them, in accordance with their own special, and varied, interests in the early thinkers. (For analyses of the doxographic tradition, and the influence of Aristotle and Theophrastus on later sources, see Mansfeld 1999, Runia 2008, and Mansfeld and Runia 1997, 2009a, and 2009b.) Although any account of a Presocratic thinker has to be a reconstruction, we should not be overly pessimistic about the possibility of reaching a historically responsible understanding of these early Greek thinkers.

Calling this group "Presocratic philosophers" raises certain difficulties. The term, coined in the eighteenth century, was made current by Hermann Diels in the nineteenth, and was meant to mark a contrast between Socrates who was interested in moral problems, and his predecessors, who were supposed to be primarily concerned with cosmological and physical speculation. "Presocratic," if taken strictly as a chronological term, is not accurate, for the last of them were contemporaneous with Socrates and even Plato. Moreover, several of the early Greek thinkers explored questions about ethics and the best way to live a human life. The term may also suggest that these thinkers are somehow inferior to Socrates and Plato, of interest only as their predecessors, and its suggestion of archaism may imply that philosophy only becomes interesting when we arrive at the classical period of Plato and Aristotle. Some scholars now deliberately avoid the term, but if we take it to refer to the early Greek thinkers who were not influenced by the views of Socrates, whether his predecessors or contemporaries, there is probably no harm in using it. (For discussions of the notion of Presocratic philosophy, see Long's introduction in Long 1999, Laks 2006, and the articles in Laks and Louguet 2002.)

A second problem lies in referring to these thinkers as philosophers. That is almost certainly not how they could have described themselves. While it is true that Heraclitus says that "those who are lovers of wisdom must be inquirers into many things" (22B35), the word he uses, *philosophos*, does not have the special sense that it acquires in the works of Plato and Aristotle, when the philosopher is contrasted with both the ordinary person and other experts, including the sophist (particularly in Plato), or in the resulting modern sense in which we can distinguish philosophy from physics or psychology; yet the Presocratics certainly saw themselves as set apart from ordinary people and also from others (certain of the poets and historical writers, for example, as we can see from Xenophanes and Heraclitus) who were their predecessors and contemporaries. As the fragment from Heraclitus shows, the early Greek philosophers thought of

themselves as inquirers into many things, and the range of their inquiry was vast. They had views about the nature of the world, and these views encompass what we today call physics, chemistry, geology, meteorology, astronomy, embryology, and psychology (and other areas of natural inquiry), as well as theology, metaphysics, epistemology, and ethics. In the earliest of the Presocratics, the Milesians, it can indeed be difficult to discern the strictly philosophical aspects of the views in the evidence available to us. Nevertheless, despite the danger of misunderstanding and thus underestimating these thinkers because of anachronism, there is an important sense in which it is quite reasonable to refer to them as philosophers. That sense is inherent in Aristotle's view (see, e.g., *Metaphysics* I, *Physics* I, *De Anima* I, *On Generation and Corruption* I): these thinkers were his predecessors in a particular sort of inquiry, and even though Aristotle thinks that they were all, for one reason or another, unsuccessful and even amateurish, he sees in them a similarity such that he can trace a line of continuity of both subject and method from their work to his own. The questions that the early Greek philosophers asked, the sorts of answers that they gave, and the views that they had of their own inquiries were the foundation for the development of philosophy as it came to be defined in the work of Plato and Aristotle and their successors. Perhaps the fundamental characteristic is the commitment to explain the world in terms of its own inherent principles.

By contrast, consider the 7th century BCE poem of Hesiod, his *Theogony* (genealogy of the gods). Hesiod tells the traditional story of the Olympian gods, beginning with Chaos, a vague divine primordial entity or condition. From Chaos, a sequence of gods is generated, often by sexual congress, but sometimes no cause for their coming to be is given. The divine figures that thus arise are often connected with a part of the physical universe, or with some aspect of human experience, so his theogony is also a cosmogony (an account of the generation of the world). The divinities (and the associated parts of the world) come to be and struggle violently

among themselves; finally Zeus triumphs and establishes and maintains an order of power among the others. Hesiod's world is one in which the major divinities are individuals who behave like super-human beings (Gaia or earth, Ouranos or sky, Cronos — an unlocated regal power, Zeus); some of the others are personified characteristics (e.g., Momus, blame; and Dusnomia, lawlessness). For the Greeks, the fundamental properties of divinity are immortality (they are not subject to death) and great power (as part of the cosmos or in managing events), and each of Hesiod's characters has these properties (even though in the story some are defeated, and seem to be destroyed). Hesiod's story is like a vast Hollywood-style family history, with envy, rage, love, and lust all playing important parts in the coming-to-be of the world as we know it. The earliest rulers of the universe are violently overthrown by their offspring (Ouranos is overthrown by Cronos, Cronos by Zeus). Zeus insures his continued power by swallowing his first consort Metis (counsel or wisdom); by this he prevents the predicted birth of rivals and acquires her attribute of wisdom (*Theogony* 886–900). In a second poem, *Works and Days*, Hesiod pays more attention to human beings, telling the story of earlier, greater creatures who died out or were destroyed by themselves or Zeus. Humans were created by Zeus, are under his power, and are subject to his judgment and to divine intervention for either good or ill. (A good discussion of the Hesiodic myths in relation to Presocratic philosophy can be found in McKirahan 2011. Burkert 2008 surveys influence from the east on the development of Presocratic philosophy, especially the myths, astronomy, and cosmogony of the Babylonians, Persians, and Egyptians.)

Hesiod's world, like Homer's, is one that is god-saturated, where the gods may intervene in all aspects of the world, from the weather to mundane particulars of human life, acting on the ordinary world order, in a way that humans, limited as they are by time, location, and narrow powers of perception, must accept but cannot ultimately understand. The Presocratics reject this account, instead seeing the world as a *kosmos*, an ordered

natural arrangement that is inherently intelligible and not subject to supra-natural intervention. A striking example is Xenophanes 21B32: "And she whom they call Iris, this too is by nature cloud / purple, red, and greeny yellow to behold." Iris, the rainbow, traditional messenger of the gods, is after all, not supra-natural, not a sign from the gods on Olympus who are outside of and immune from the usual world order; rather it is, in its essence, colored cloud.

Calling the Presocratics *philosophers* also suggests that they share a certain outlook with one another; an outlook that can be contrasted with that of other early Greeks. Although scholars disagree about the extent of the divergence between the early Greek philosophers and their non-philosophical predecessors and contemporaries, it is evident that Presocratic thought exhibits a difference not only in its understanding of the nature of the world, but also in its view of the sort of explanation of it that is possible. This is clear in Heraclitus. Although Heraclitus asserts that those who love wisdom must be inquirers into many things, inquiry alone is not sufficient. At 22B40 he rebukes four of his predecessors: "Much learning does not teach understanding; else it would have taught Hesiod and Pythagoras, and again Xenophanes and Hecataeus." Heraclitus' implicit contrast is with himself; in 22B1 he suggests that he alone truly understands all things, because he grasps the account that enables him to "distinguish each thing in accordance with its nature" and say how it is. For Heraclitus there is an underlying principle that unites and explains everything. It is this that others have failed to see and understand. According to Heraclitus, the four have amassed a great deal of information — Hesiod was a traditional source of information about the gods, Pythagoras was renowned for his learning and especially views about how one ought to live, Xenophanes taught about the proper view of the gods and the natural world, Hecataeus was an early historian — but because they have failed to grasp the deeper significance of the facts available to them, their unconnected bits of knowledge do not constitute

understanding. Just as the world is a *kosmos*, an ordered arrangement, so human knowledge of that world must be ordered in a certain way.

2. The Milesians

In his account of his predecessors' searches for "causes and principles" of the natural world and natural phenomena, Aristotle says that Thales of Miletus (a city in Ionia, on the west coast of what is now Turkey) was the first to engage in such inquiry. He seems to have lived around the beginning of the 6th c. BCE. Aristotle mentions that some people, before Thales, placed great importance on water, but he credits Thales with declaring water to be the first cause (*Metaphysics* 983b27–33), and he then later raises the question of whether perhaps Hesiod was the first to look for a cause of motion and change (984b23ff.). These suggestions are rhetorical: Aristotle does not seriously imply that those he mentions are engaged in the same sort of inquiry as he thinks Thales was. Two other Greek thinkers from this very early period, Anaximander and Anaximenes, were also from Miletus, and although the ancient tradition that the three were related as master and pupil may not be correct, there are enough fundamental similarities in their views to justify treating them together.

The tradition claims that Thales predicted a solar eclipse in 585 BC (11A5), introduced geometry into Greece from Egypt (11A11), and produced some engineering marvels. Anaximander is reported to have invented the gnomon (the raised piece of a sundial whose shadow marks time); to have created a sphere of the heavens serving as an astronomical and cosmological model (12A1); and to have been the first to draw a map of the inhabited world (12A6). Regardless of whether these reports are correct (and in the case of Thales' prediction they almost certainly are not), they indicate something important about the Milesians: their interests in measuring and explaining celestial and terrestrial phenomena were as

strong as their concern with the more abstract inquiries into the causes and principles of substance and change attributed to them by Aristotle (Algra 1999, White 2002 and 2008). They did not see so-called "scientific" and "philosophical" questions as belonging to separate disciplines, requiring distinct methods of inquiry. The assumptions and principles that we (along with Aristotle) see as constituting the philosophical foundations of their theories are, for the most part, implicit in the claims that they make. Nevertheless, it is legitimate to treat the Milesians as having philosophical views, even though no clear statements of these views or specific arguments for them can be found in the surviving fragments and testimonia.

Aristotle's comments do not sound as if they were based on first-hand knowledge of Thales' views, and the doxographical reports say that Thales did not write a book. Yet Aristotle is confident that Thales belongs, even if honorifically, to that group of thinkers that he calls "inquirers into nature" and distinguishes him from earlier poetical "myth-makers." In Book I of the *Metaphysics*, Aristotle claims that the earliest of these, among whom he places the Milesians, explained things only in terms of their matter (*Met*. I.3 983b6–18). This claim is anachronistic in that it presupposes Aristotle's own novel view that a complete explanation must encompass four factors: what he called the material, efficient, formal, and final causes. Yet there is something in what Aristotle says. Aristotle links Thales' claim that the world rests on water with the view that water was the *archē*, or fundamental principle, and he adds that "that from which they come to be is a principle of all things" (983b24–25; 11A12). He suggests that Thales chose water because of its fundamental role in coming-to-be, nutrition, and growth, and claims that water is the origin of the nature of moist things.

Aristotle's general assertion about the first thinkers who gave accounts of nature (and his specific discussion of Thales' reliance on water as a first

principle) brings out a difficulty in interpreting the early Presocratics. According to Aristotle's general account, the Presocratics claimed that there was a single enduring material stuff that is both the origin of all things and their continuing nature. Thus, on this view, when Thales says that the first principle is water, he should be understood as claiming both that the original state of things was water and that even now (despite appearances), everything is really water in some state or another. The change from the original state to the present one involves changes in the material stuff such that although it may not now appear to be water everywhere (but seems to be airier or earthier than water in its usual state, or its original one), there is no transformation of water into a different kind of stuff (air or earth, for instance). Yet, when Aristotle comes to give what details he can of Thales' view, he suggests only that for Thales, water was the first principle because everything comes from water. Water, then, was perhaps the original state of things for Thales, and water is a necessary condition for everything that is generated naturally, but Aristotle's summary of Thales' view does not imply that Thales claimed that water endures through whatever changes have occurred since the original state, and now just has some new or additional properties. Thales may well have thought that certain characteristics of the original water persisted: in particular its capacity for motion (which must have been innate in order to generate the changes from the original state). This is suggested by Thales' reported claims that the lodestone (with its magnetic properties) and amber (which when rubbed exhibits powers of attraction through static electricity) have souls and that all things are full of gods. Aristotle surmises that Thales identified soul (that which makes a thing alive and thus capable of motion) with something in the whole universe, and so supposed that everything was full of gods (11A22)—water, or soul, being a divine natural principle. Certainly the claim that the lodestone has soul suggests this account. Given that the analysis of change (both qualitative and substantial) in terms of a substratum that gains and loses properties is

Aristotelian (although perhaps foreshadowed in Plato), it is not surprising that the earlier views were unclear on this issue, and it is probable that the Milesian view did not clearly distinguish the notions of an original matter and an enduring underlying stuff (Graham 2006).

The reports about Thales show him employing a certain kind of explanation: ultimately the explanation of why things are as they are is grounded in water as the basic stuff of the universe and the changes that it undergoes through its own inherent nature. In this, Thales marks a radical change from all other previous sorts of accounts of the world (both Greek and non-Greek). Like the other Presocratics, Thales sees nature as a complete and self-ordering system, and sees no reason to call on divine intervention from outside the natural world to supplement his account—water itself may be divine, but it is not something that intervenes in the natural world from outside (Gregory, 2013). While the evidence for Thales' naturalistic account is circumstantial, this attitude can be directly verified for Anaximander.

In the one fragment that can be securely attributed to Anaximander (although the extent of the implied quotation is uncertain), he emphasizes the orderly nature of the universe, and indicates that the order is internal rather than imposed from outside. Simplicius, a 6th c. CE commentator on Aristotle's *Physics*, writes:

> Of those who say that [the first principle] is one and moving and indefinite, Anaximander, son of Praxiades, a Milesian who became successor and pupil to Thales, said that the indefinite (*to apeiron*) is both principle (*archē*) and element (*stoicheion*) of the things that are, and he was the first to introduce this name of the principle. He says that it is neither water nor any other of the so-called elements, but some other indefinite (*apeiron*) nature, from which come to be all the heavens and the worlds in them; and those things, from

which there is coming-to-be for the things that are, are also those into which is their passing-away, in accordance with what must be. For they give penalty (*dikê*) and recompense to one another for their injustice (*adikia*) in accordance with the ordering of time—speaking of them in rather poetical terms. It is clear that having seen the change of the four elements into each other, he did not think it fit to make some one of these underlying subject, but something else, apart from these. (Simplicius, *Commentary on Aristotle's Physics* 24, lines 13ff. = 12A9 and B1)[1]

Thus, there is an original (and originating) indefinite stuff, from which all the heavens and the worlds in them come to be. This claim probably means that the original state of the universe was an indefinitely large mass of stuff that was also indefinite in its character.[2] This stuff then gave rise through its own inherent power to the ingredients that themselves constitute the world as we perceive it.

A testimony about Anaximander from Pseudo-Plutarch (12A10) says that "Something productive of hot and cold was separated off from the eternal at the genesis of this world and from this a sphere of flame grew around the air around the earth like the bark around a tree." Neither the cause nor the precise process of separation is explained, but it is probable that Anaximander would have thought of motion as innate and so that the original source of change was part of the character of the indefinite itself. The passage from Simplicius shows that Anaximander does not think that the eternal indefinite stuff gives rise directly to the cosmos as we know it. Rather, relying on a semi-biological model, Anaximander claims that the *apeiron* somehow generates the opposites hot and cold. Hot and cold are themselves stuffs with powers; and it is the actions of these stuffs/powers that produce the things that come to be in our world. The opposites act on, dominate, and contain each other, producing a regulated structure; thus things pass away into those things from which they came to be. It is this

structured arrangement that Anaximander refers to when he speaks of justice and reparation. Over the course of time, the cycles of the seasons, the rotations of the heavens, and other sorts of cyclical change (including coming-to-be and passing-away) are regulated and thus form a system. This system, ruled by the justice of the ordering of time is in sharp contrast with the chaotic and capricious world of the personified Greek gods who interfere in the workings of the heavens and in the affairs of human beings (Kahn 1985a, Vlastos 1947, Guthrie 1962).

The pattern that can be seen in Thales and Anaximander of an original stuff giving rise to the phenomena of the cosmos continues in the views of the third of the Milesians, Anaximenes. He replaces Anaximander's *apeiron* with air, thus eliminating the first stage of the coming-to-be of the cosmos (the something productive of hot and cold). Rather, he returns to an originating stuff more like Thales' water. In 13A5, Aristotle's associate Theophrastus, quoted by Simplicius, speculates that Anaximenes chose air because he agreed that a basic principle must be neutral (as Anaximander's *apeiron* is) but not so lacking in properties that it seems to be nothing at all. Air can apparently take on various properties of color, temperature, humidity, motion, taste, and smell. Moreover, according to Theophrastus, Anaximenes explicitly states the natural mechanism for change; it is the condensation and rarefaction of air that naturally determine the particular characters of the things produced from the originating stuff. Rarified, air becomes fire; more and more condensed, it becomes progressively wind, cloud, water, earth, and finally stones. "The rest," says Theophrastus, "come to be from these." Plutarch says that condensation and rarefaction are connected with cooling and heating, and he gives the example of breath (13B1). Releasing air from the mouth with compressed lips produces cool air (as in cooling soup by blowing on it), but relaxed lips produce warm air (as when one blows on cold hands to warm them up).

Does the originating stuff persist through the changes that it undergoes in the generating processes? Aristotle's account suggests that it does, that Anaximenes, for instance, would have thought that stone was really air, although in an altered state, just as we might say that ice is really water, cooled to a point where it goes from a liquid to a solid state. Because the water does not cease to be water when it is cooled and becomes ice, it can return to a liquid when heated and then become a gas when more heat is applied. On this view, the Milesians were material monists, committed to the reality of a single material stuff that undergoes many alterations but persists through the changes (Barnes 1979, Guthrie 1962, Sedley 2007 and 2009). Yet there are reasons to doubt that this was actually the Milesian view. It presumes that the early Greek thinkers anticipated Aristotle's general theory that change requires enduring underlying substances that gain and lose properties. The earliest Greeks thought more in terms of powers (Vlastos 1947, Heidel 1906), and the metaphysical problem of what it is to be a substance was yet to be formulated. Clearly the Milesians were interested in the originating stuff from which the world developed (Anaximander and Anaximenes are explicit about transformations of such an eternal originating stuff), but the view that this endured as a single substratum may not have been theirs. Rather, it has been suggested by Graham (1997 and 2006; Mourelatos 2008) that the Milesians were not, in Aristotle's sense, material monists. On this view, the original/originating stuff is transformed into other substances. Anaximenes, for instance, may have thought that the change from air to water does not involve the persistence of air as any sort of substratum. There is no special role that air plays in the theory except that it is the originating stuff and so first in an analysis of the law-like cyclical changes that produce various stuffs as the cosmos develops (Graham 2006, ch. 4). Such an interpretation suggests how different the Milesian conception of the world is from Aristotle's.

3. Xenophanes of Colophon and Heraclitus of

Ephesus

Living in the last years of the 6th c. and the beginning of the 5th, Xenophanes and Heraclitus continue the Milesian interest in the nature of the physical world, and both offer cosmological accounts; yet they go further than the Milesians not only through their focus on the human subject and the expanded range of their physical explanations, but by investigating the nature of inquiry itself. Both explore the possibility of human understanding and question its limits. Recent work on Xenophanes' epistemology and his cosmology has made much of his scientific work clearer and more impressive (Lesher 1992, Mourelatos 2008). He has, to a great extent, been rescued from his traditional status as a minor traveling poet-sage who railed against the glorification of athletes and made some interesting comments about the relativity of human conceptions of the gods. Instead, he has come to be seen as an original thinker in his own right who influenced later philosophers trying to characterize the realms of the human and the divine, and exploring the possibility that human beings can gain genuine knowledge and wisdom, i.e., are able to have a god's eye view of things and understand them (Curd 2013, Mogyoródi 2002 and 2006).

Xenophanes claims that all meteorological phenomena are clouds, colored, moving, incandescent: rainbow, St. Elmo's Fire, the sun, the moon. Clouds are fed by exhalations from the land and sea (mixtures of earth and water). The motions of earth and water, and hence of clouds, account for all the things we find around us. His explanations of meteorological and heavenly phenomena lead to a naturalistic science:

> She whom they call Iris, this too is by nature (*pephuke*) cloud
> purple, and red, and greeny-yellow to behold. (21B32)

Xenophanes says that the star-like phenomena seen when aboard ship, which some call the Dioscuri, are cloudlets, glimmering because of their kind of motion. (A39)

In the 1980's Alexander Mourelatos argued that Xenophanes employs an important new pattern of explanation: *X* is really *Y*, where *Y* reveals the true character of *X*. Xenophanes signals this by the use of *pephuke* in B32, and no doubt it (or some word like it) was there in the original of A39 as well. Xenophanes thus provides an account of a phenomenon often taken to be a sign from the divine—Iris as the messenger; the Dioscuri (St. Elmo's fire) as comfort for sailors—that reduces it to a natural occurrence.

That meteorological phenomena are not divine is not all that Xenophanes has to say about the gods. He notes anthropomorphic tendencies in conceptions of the gods (B14: "Mortals suppose that the gods are born, and have their own dress, voice, and body;" B16: "Ethiopians say that their gods are snub-nosed and dark, Thracians, that theirs are grey-eyed and red-haired"). He also famously suggests that horses, oxen, and lions would have equine, bovine, and leonine gods (B15). Yet Xenophanes also makes positive claims about the nature of the divine, including the claim that there is a single greatest god:

One god greatest among gods and men,
Resembling mortals neither in body nor in thought.
… whole [he] sees, whole [he] thinks, and whole [he] hears,
but completely without toil he agitates all things by the
    thought of his mind.
… always he remains in the same (state), agitated not at all,
nor is it fitting that he come and go to different places at different
times. (B23, 24, 25, 26)

While indifferent to the affairs of human beings, Xenophanes' divine being comprehends and controls a cosmos that is infused with thinking: it is understood, organized, and managed by divine intellection. Having removed the gods as bearers of knowledge to humans, and denied that the divine takes an active interest in what mortals can or cannot know, Xenophanes asserts the conclusion to be drawn from his naturalistic interpretation of phenomena: the gods are not going to reveal anything to us; we are epistemologically autonomous and must rely on our own capacity for inquiry. That way, we "discover better," as he says in B18, a fragment that is optimistic about the capacities of human intelligence (see Lesher 1991):

> Indeed not even from the beginning did the gods indicate all things
> to mortals, but, in time, inquiring, they discover better.

This suggests that human thought can mimic divine understanding, at least to some degree. Xenophanes' own practice seems consistent with the claims of B18; his own inquiries and explanations led him to unified explanations of terrestrial and celestial phenomena. Yet B34 suggests skepticism:

> And of course the clear and certain truth no man has seen,
> nor will there be anyone who knows about the gods and what I say
> about all things;
> for even if, in the best case, he should chance to speak what is the
> case,
> all the same, he himself does not know; but opinion is found over
> all.

Whether this is global or limited skepticism is controversial (Lesher 1992 and 1994 argues for a limited interpretation). Xenophanes stresses the difficulty of coming to certainty, particularly about things beyond our

direct experience. Nonetheless, in B35 (a tantalizingly short fragment), Xenophanes says, "Let these thing be accepted to be like the truth" (see Bryan 2012 for a full discussion).

Famously obscure, accused by Plato of incoherence and by Aristotle of denying the law of non-contradiction, Heraclitus writes in an aphoristic style. His apparently paradoxical claims present difficulties to any interpreter. Nevertheless, he raises important questions about knowledge and the nature of the world. The opening of Heraclitus' book refers to a "*logos* which holds forever."[3] There is disagreement about exactly what Heraclitus meant by using the term *logos*, but it is clear from 22B1 and B2 as well as B50 and other fragments that he refers to an objective law-like principle that governs the cosmos, and which it is possible (but difficult) for humans to come to understand. There is a single order that directs all things ("all things are one" B50); this order is divine, and is sometimes connected by humans with the traditional gods (it is "both unwilling and willing to be called by the name of Zeus" B32). Just as Zeus, in the traditional view, controls everything from Olympus with a thunderbolt, so this single ordered system also steers and controls the whole cosmos, but from within. The sign of the unchanging order of the eternal system is fire —just as fire is always changing and always the same, the *logos*, itself permanent, contains the unchanging account that explains the alterations and transformations of the cosmos.

This plan or order that steers the cosmos is, itself, a rational order. This means not only that it is non-capricious and so intelligible (in the sense that humans can, at least in principle, come to understand it), it is also an intelligent system: there is an intelligent plan at work, if only in the sense of the cosmos working itself out in accordance with rational principles.[4] Consider B114:

> Those who would speak with understanding must ground

themselves firmly in that which is common to all, just as a city does in its law, and even more firmly! For all human laws are nourished by one law, the divine; for it rules as far at it wishes and suffices for all, and is still more than enough.

Heraclitus is not only claiming that human prescriptive law must harmonize with divine law, but he is also asserting that divine law encompasses both the universal laws of the cosmos itself and the particular laws of humans. The cosmos itself is an intelligent, eternal (and hence divine) system that orders and regulates itself in an intelligent way: the *logos* is the account of this self-regulation. We can come to grasp and understand at least part of this divine system. This is not merely because we ourselves are part of (contained in) the system, but because we have, through our capacity for intelligent thinking, the power to grasp the system as a whole, through knowing the *logos*. How this grasping is supposed to work is tantalizingly obscure.

Heraclitus regards the cosmos as an ordered system like a language that can be read or heard and understood by those who are attuned to it. That language is not just the physical evidence around us ("Eyes and ears are bad witnesses to those with barbarian souls" B107); the sheer accumulation of information is not the same as wisdom (see the rebuke in 22B40, quoted above). Although the evidence of the senses is important (see B55 and other fragments on direct experience vs. hearsay), careful and thoughtful inquiry is also necessary. Those who are lovers of wisdom must be good inquirers into many things (B35; also B101: "I enquired into myself"), and must be able to grasp how the phenomena are signs or evidence of the larger order; as Heraclitus notes in B123, "nature is accustomed to hide itself," and the evidence must be interpreted carefully. That evidence is the interplay of opposing states and forces, which Heraclitus points to by claims about the unity of opposites and the roles of strife in human life as well as in the cosmos. There are fragments that

proclaim the unity or identity of opposites: the road up and down are one and the same (B60), the path of writing is both straight and crooked (B59), sea water is very pure and very foul (B61). The famous river fragments (B49a, B12, B91a) question the identity of things over time, while a number of fragments point to the relativity of value judgments (B9, B82, B102). Anaximander's orderly arrangement of just reciprocity governed by time is replaced by a system ruled by what Heraclitus calls war: "It is right to know that war is common and justice strife, and that all things come to be through strife and are so ordained" (B80). This strife or war is the set of changes and alterations that constitute the processes of the cosmos. These changes are regular and capable of being understood by one who can speak the language of the *logos* and thus interpret it properly (see Long, 2009). Although the evidence is confusing, it points to the deeper regularities that constitute the cosmos, just as Heraclitus' own remarks can seem obscure yet point to the truth. Heraclitus surely has his own message (and his delivery of it) in mind in B93, "The lord whose oracle is at Delphi neither speaks nor conceals, but gives a sign."

One of the earliest of the Greek philosophers to discuss the human soul, Heraclitus' claims about it, like his other views, are expressed enigmatically. Yet it seems fairly clear that he treats soul as the seat of emotion, movement, and intellect. B107 (quoted above) indicates that understanding is a function of soul, and in B117, the drunken man who must be led by a boy because he has lost control of his legs, and also does not know where he goes or what he does. Drunkenness is the cause of all this: because his soul has become wet its powers are dampened down and become ineffective. B118 asserts "gleam of light: dry soul, wisest and best." This suggests that for Heraclitus, soul is a stuff that is affected by changes along the hot/cold and wet/dry continua (the gleam of light suggests a fiery, i.e., hotter soul is best). Indeed in B36, soul is listed as one of the stages of transformation of the cosmic stuffs: "it is death to souls to become water, and to water death to become earth; from earth

water comes to be, from from water, soul." Although Heraclitus says that it is only divine nature that has complete understanding (B78), his linking of fire with the *logos* and the divine, along with his view that the best and wisest soul is hot and dry, suggests that humans who care for their souls and search for the truth contained in the *logos* can overcome human ignorance and approach the understanding that Heraclitus himself has obtained. (Betegh 2007, 2009, 2013 and Dilcher 1995 both discuss the nature and importance of soul for Heraclitus; see also Granger 2000 and Kahn 1979.)

4. Parmenides of Elea

Parmenides, born ca. 510 BCE in the Greek colony of Elea in southern Italy (south of Naples, and now known as Velia), explores the nature of philosophical inquiry, concentrating less on knowledge or understanding (although he has views about these) than on what can be understood. Xenophanes identified genuine knowledge with the grasping of the sure and certain truth and claimed that "no man has seen" it, at least with respect to some topics (21B34); Heraclitus had asserted that divine nature, not human, has right understanding (22B78), although he implies that some humans can acquire divine-like understanding. Parmenides argues that human thought can reach genuine knowledge or understanding, and that there are certain marks or signs that act as guarantees that the goal of knowledge has been reached. A fundamental part of Parmenides' claim is that what *must* be (cannot not-be, as Parmenides puts it) is more knowable than what is merely contingent (what may or may not be), which can be the object only of belief.

Parmenides gives us a poem in Homeric hexameters, narrating the journey of a young man (a *kouros*, in Greek) who is taken to meet a goddess who promises to teach him "all things" (28B1). The content of the story the goddess tells is not the knowledge that will allow humans, by having it, to

know. Rather, the goddess gives the *kouros* the tools to acquire that knowledge himself:

> It is right that you learn all things,
> Both the unshaking heart of well-persuasive truth,
> and the beliefs of mortals, in which there is no true trust.
> But nevertheless, you shall learn these things too, how it were right
> that the things that seem be reliably, being indeed the whole of
> things. (B1.28–32)

The goddess does not provide the *kouros* with a list of true propositions, as a body of knowledge for him to acquire, and false ones to be avoided. Rather, in teaching him how to evaluate claims about what-is, the goddess unleashes the *kouros'* own cognitive powers to know everything, by testing and evaluation, accepting or rejecting claims about the ultimate nature of things— for that alone is capable of being known. For Parmenides, the mark of what is known is that it is something that genuinely is, with no taint of what-is-not. That is why, for him, what-is not only is, but must be and cannot not-be. He sets this out in the key passages of B2 and B3:[5]

> Come now, and I will tell you, and you, hearing, preserve the story,
> the only routes of inquiry there are for thinking;
> the one that it is and that it cannot not be
> is the path of Persuasion (for it attends upon truth)
> the other, that it is not and that it is right that it not be,
> this I point out to you is a path wholly inscrutable
> for you could not know what is not (for it is not to be
> accomplished)
> nor could you point it out… For the same thing is for thinking and
> for being.

The routes are methods of inquiry: keeping on the correct route will bring one to what-is, the real object of thought and understanding. Although what the goddess tells the *kouros* has divine sanction (hers), that is not why he should accept it. Rather, the truth she tells reveals a mark of its own truth: it is testable by reason or thought itself. In B7 the goddess warns that we must control our thought in the face of the ever-present seductions of sense-experience:

> For never shall this be forced through: that things that are not are;
> but restrain your thought from this route of inquiry,
> nor let much-experienced habit force you along this path,
> to ply an aimless eye and resounding ear
> and tongue, but judge by reasoning (*logos*) the much-battled testing
> spoken by me.

The *kouros* himself can reach a decision or determination of the truth solely through use of his *logos*. *Logos* here means thinking or reasoning (Parmenides probably means the human capacity for thought in general). The test (restated at B8.15–16), is "is or is not?" This is not just a question of non-contradiction (which would give us coherence), but an inquiry whether or not the supposition that something is would entail, on further examination, the reality of what-is-not (which is impossible).

The arguments of B8 demonstrate *how* what-is must be. In applying these arguments as tests against any suggested basic entity in the Presocratic search for ultimate causes or principles, the *kouros* can determine whether or not a proposed theory is acceptable. For Parmenides *noos* is not itself an infallible capacity. One can think well or badly; correct thinking is that which takes the correct route and so reaches what-is. The mortals on the incorrect route are thinking, but their thoughts have no real object (none that is real in the appropriate way), and so cannot be completed or

perfected by reaching the truth. In B8 Parmenides sets out the criteria for the being of what-is, and then the arguments for those criteria:

> … a single account still
> remains of the route that it is; and on this route there are
> very many signs, that what-is is ungenerable and imperishable,
> a whole of a single kind, and unshaking and complete;
> nor was it nor will it be, since it is now all together
> one, cohesive. (B8.1–6)

Any thing that genuinely is cannot be subject to coming-to-be or passing-away, must be of a single nature, and must be complete, in the sense of being unchangeably and unalterably what it is. These are signs for what any ultimate cause or principle must be like, if it is to be satisfactory as a principle, as something that can be known. The signs are adverbial, showing how what-is is (Mourelatos 2008). Only an entity which is in the complete way can be grasped and understood in its entirety by thought. McKirahan (2008) provides a thorough analysis of the arguments of B8, as does Palmer (2009).

After laying out the arguments about what-is, the goddess turns to the route of mortals, in an account which she calls "deceptive." Although Parmenides has been read as thus rejecting any possibility of cosmological inquiry (Barnes 1979, Owen 1960), there are persuasive interpretations that allow for justified belief about the contingent world, a world that may or may not be, and is not such that it must be (Nehamas 2002, Curd 2004, Palmer 2009). The problem of mortals is that they mistake what they perceive for what there is (and must be). As long as one realizes that the world of perception is not genuinely real, and cannot therefore be the object of knowledge, it may be possible for there to be justified belief about the cosmos. Some details of Parmenides' own cosmology are given, arguably as justified belief, in the *Doxa* section of the poem, and more in

the testimonia from later authors. Parmenides seems to have been the first Presocratic to claim that the moon gets its light from from the sun and that the earth is spherical. Recently scholars have focused on these claims about the natural world, and have argued that Parmenides should be understood as offering an account of appearances that can and should be deemed acceptable (Palmer 2009, Cordero 2010, Graham 2013, Mourelatos 2013, Bryan 2012, Johansen forthcoming). Nevertheless, Parmenides marks a sharp distinction between being (what-is and must be) and becoming, and between knowledge and perception-based belief or opinion.

5. The Pythagorean Tradition

In the last quarter of the sixth century, before Parmenides' birth, Pythagoras of Samos (an Aegean island) arrived in Croton, in southern Italy. He established a community of followers who adopted his political views, which favored rule by the "better people," and also the way of life he recommended on what seem to have been more or less philosophical bases. The traditional view has been that the aristocracy, the "better people," generally meant the rich. But Burkert notes that as early as the 4th c. BCE there were two traditions about Pythagoras, one that meshes with the traditional view and associates Pythagoras with political tyrants, and another that credits him with rejecting tyrannies for aristocracies that might not have been grounded in wealth (Burkert 1972, 119). The Pythagorean Archytas (born late 5th century) lived in a democracy (Tarentum in southern Italy), and seems to have argued for fair and proportionate dealings between rich and poor (Huffman 2005). The Pythagorean way of life included adherence to certain prescriptions including religious rites and dietary restrictions (there is a general discussion in Kahn 2001). Detailed treatment of Pythagoras and

Pythagoreanism can be found in Zhmud (2012 and 2013); an excellent collection of articles on Pythagoreanism is in Huffman (ed.) 2014.

Like Socrates, Pythagoras wrote nothing himself, but had a great influence on those who followed him. He had a reputation for great learning and wisdom (see Empedocles 31B129), although he was treated satirically by both Xenophanes (21B7) and Heraclitus (22B40, B129). We do not know to what extent this included knowledge of mathematics, as would be suggested by the attribution to him of the famous Pythagorean theorem of geometry (Rowett 2013). The details of Pythagoras' views are unclear, but he seems to have advocated the reincarnation of the soul (a novel idea among the Greeks, also developed in Orphic religion) and the possibility of the transmigration of the human soul after death into other animal forms. Pythagorean writers after his own time stressed the mathematical structure and order of the universe. This is often attributed directly to Pythagoras (primarily because of the geometrical theorem that bears his name), but recent scholarship has shown that the evidence for attributing this mathematically-based cosmology to Pythagoras himself is convoluted and doubtful (Burkert 1972, Huffman 1993 and 2005; but see Zhmud 1997).

What seems clear is that the early Pythagoreans conceived of nature as a structured system ordered by number (see the SEP entry on Pythagoras), and that such post-Parmenidean Pythagoreans as Philolaus (last half of the 5th century, more than a generation after Pythagoras' death) and Archytas (late 5th to early 4th century) held more complicated views about the relation between mathematics and cosmology than it is reasonable to suppose Pythagoras himself could have advanced. The Pythagorean tradition thus includes two strains. There are reports of a split in the period after Pythagoras' death between what we would term the more philosophically inclined Pythagoreans and others who primarily adopted the Pythagorean ethical, religious and political attitudes. The latter, called

the *acusmatici*, followed the Pythagorean precepts, or *acusmata* (which means "things heard"). The former, the philosophical Pythagoreans (including Philolaus and Archytas), were the so-called *mathematici*, and while they recognized that the *acusmatici* were indeed Pythagoreans by virtue of accepting Pythagorean precepts, they claimed that they themselves were the true followers of Pythagoras.

Philolaus of Croton seems to have blended the Pythagorean life with an awareness of and appreciation for the arguments of Parmenides (Huffman 1993). According to Philolaus, "Nature in the cosmos was fitted together out of unlimiteds and limiters" (44B1). These limiters and unlimiteds play the role of Parmenidean basic realities—they are and unchangingly must be what they are, and so can be known; they are joined together in a harmonia (literally, a carpenter's joint; metaphorically, a harmony), and "it was not possible for any of the things that are and are known by us to come to be, without the existence of the being of things from which the cosmos was put together" (44B6). The unlimiteds are unstructured stuffs and continua; the limiters impose structure (shape, form, mathematical structure) on the unlimiteds. Things become knowable because they are structured in this way; the structure can apparently be expressed in a numerical ratio that allows for understanding: "All things that are known have number; for without this nothing whatever could possibly be thought of or known" (44B4). Philolaus also developed a theory of the cosmos that displaced the earth from the center, replaced by what he called Hestia, the central fire (Graham 2013, 2014), and offered novel accounts of eclipses.

6. Other Eleatics: Zeno and Melissus

Parmenides had argued that there were strict metaphysical requirements on any object of knowledge; the later Eleatics (named for following Parmenidean doctrines rather than for strictly geographical reasons), Zeno of Elea (born ca. 490) and Melissus of Samos (fl. ca. 440), extended and

explored the consequences of his arguments. Zeno paid particular attention to the contrast between the requirements of logical argument and the evidence of the senses (Vlastos 1967 is a masterly treatment of Zeno; see also McKirahan 1999 and 2005). The four famous paradoxes of motion, for which he is now and in antiquity best known, purported to show that, despite the evidence all around us, the ordinary motion of everyday experience is impossible. The paradoxes claim that motions can never be begun (the Achilles) or be completed (the Dichotomy), entail contradictions (the Moving Blocks), or are altogether impossible (the Arrow).[6] Recent philosophers of space and time (see Grünbaum 1967, articles in Salmon 2001, Huggett 1999, SEP entry on Zeno's Paradoxes) hold that the arguments are reductios of the theses that space and time are continuous (the Achilles and the Dichotomy) or discrete (the Moving Blocks and the Arrow). Consider the Dichotomy: a runner can never complete a run from point A to point B. First, the runner must move from A to a point halfway between A and B (call it C). But between A and C there is yet another halfway point (D), and the runner must first reach D. But between A and D there is yet another halfway point … and so on, ad infinitum. So the runner, starting at A, can never reach B. The argument assumes that it is impossible to pass an infinite number of points in a finite time. Similarly, Zeno produced paradoxes showing that plurality is impossible: if things are many, contradictions follow (Plato's *Parmenides* 127e1ff.; Zeno in 29B1, 29B2, and 29B3); there were also purported proofs that place is impossible (29A24) and that things cannot have parts (the Millet Seed, 29A29).

Melissus, dismissed as a simple-minded thinker by Aristotle (and by some contemporary scholars as well but see Makin 2005), expands Parmenides' arguments about the nature of what-is (Palmer 2004). It is Melissus who explicitly claims that only one thing can be: if what-is is unlimited (as he thinks it is), it must be one and all alike (if there were two [in number or in character] they would be "limited against each other" 30B6). Melissus

specifically argues against the empty (the void), and rejects the possibility of rearrangement (which would allow for the appearance of coming-to-be and passing-away, and of movement)—all these characteristics are incompatible with the unity of what-is (i.e., the One). Melissus thus claims that what is real is completely unlike the world that we experience: the split between appearance and reality is complete and unbridgeable.

7. The Pluralists: Anaxagoras of Clazomenae and Empedocles of Acragas

While Zeno and Melissus reinforced Parmenides' distinction between what-is (i.e., what must be) and what appears, other post-Parmenidean thinkers accepted Parmenides' arguments against coming-to-be and passing-away (as characterizing what-is), and about the stable nature of what is ultimately real, and argued that these arguments did not rule out the possibility of metaphysically-based (or rational) cosmology. Both Anaxagoras and Empedocles worked within the Parmenidean pattern while developing distinct cosmological systems that addressed their own particular concerns (especially in the case of Empedocles, concerns about the proper way to live).

Anaxagoras (writing in the mid-5th c.) claims, "The Greeks [i.e., ordinary people] do not think correctly about coming-to-be and passing-away; for no thing comes to be or passes away, but is mixed together and dissociated from the things that are. And thus they would be correct to call coming-to-be mixing-together and passing-away dissociating" (59B17). What seem to be generated objects (human beings, plants, animals, the moon, the stars) are instead temporary mixtures of ingredients (such as earth, air, fire, water, hair, flesh, blood, dense, dark, rare, bright, and so on). Recent treatments of Anaxagoras (Marmodoro 2015) have suggested that the ingredients are primarily powers that manifest themselves in the mixtures produced.[7] The original state was universal mixture: "All things were

together, unlimited both in amount and in smallness, for the small, too, was unlimited. And because all things were together, nothing was evident" (59B1). This mixture is set into rotary motion by the operation of Mind (*Nous* – B12, B13, B14; see discussions in Laks 1993, Lesher 1995, Menn 1995, Curd 2007), a separate cosmic entity that does not share in such mixture. As the rotation spreads out through the unlimited mass of indistinguishably intermingled ingredients, the rotation causes a winnowing or separating effect, and the cosmos as we know it emerges from the mixture. Moreover, not only *were* all things together, they are *even now* all together, in a different way, despite the differentiations now achieved. Everything is in everything (59B5, B6, B11), in some proportions, however small or great – this is a move to prevent even the appearance of coming-to-be from what-is-not.

Anaxagoras marks an important theoretical step in attributing the motion of his ingredients to an independent, intelligent force (although both Plato and Aristotle were disappointed that his theory was not properly—from their point of view—teleological; on this see Sedley 2007, Curd forthcoming). The rotation begun by Mind is causally responsible for the formation of the heavens and the activities of the great masses of the earth and the water on the earth, as well as all meteorological phenomena. Insofar as the causes of the operations of the heavens and the phenomena apparent to us from day to day are the same at both the macro- and micro-level (the rotations that cause the apparent motions of the stars are the same as those that govern the cycles of weather and life and death on earth), we can infer the nature of what is real from what is apparent (Anaxagoras' scientific views are treated in Graham 2006 and 2013). Although we do not perceive all things as being together, and the move to the ultimate explanations is an inference, it is a legitimate one ("owing to their [the senses'] feebleness, we are not able to determine the truth" yet "appearances are a sight of the unseen" 59B21 and 21a).

A younger contemporary of Anaxagoras, Empedocles, who lived in Sicily, also recognized the force of Parmenides' arguments against coming-to-be and passing-away. (Empedocles also adopts Parmenides' poetic meter in order to tell his story.) Empedocles proposes a cosmos formed of the four roots (as he calls them), earth, water, air, and fire along with the motive forces of Love and Strife. It is often claimed that, for Empedocles, Love simply produces mixture and Strife only causes separation. Empedocles' view is more complicated, for both forces mix and separate. Love unites opposed (unlike) things by pulling apart and then mixing these unlikes, while Strife sets unlikes in opposition and segregates them, hence Strife mixes like with like. Just as painters can produce fantastically lifelike scenes just by mixing colors, so the operations of Love and Strife, using just the four roots can produce "trees and men and women, and beasts and birds and water-nourished fish, and long-lived gods best in honors" (31B17). These are the things that Empedocles calls "mortal," and he even provides recipes. 31B73 tells how Kypris (the goddess Aphrodite, i.e. love) fashions shapes (or kinds): "she moistened earth in rain, and gave it to quick fire to harden." B96 gives a recipe for bones, while in B98 flesh and blood have the same recipe (earth, water, air, and fire in equal proportions), but differ in the refinement of the mixture.

Like the other Presocratics, Empedocles has a cosmological theory, in his case, an unending cycle involving the competition between Love and Strife. Love overcomes the separating influence of Strife, bringing together unlikes and so preventing the clinging together of likes. The triumph of Love results in the Sphere, which is a complete mixture because the four unlike roots are as mixed (integrated) as possible. Strife breaks up the sphere by beginning to attract like to like and so pulling the mixture apart, until, when it triumphs, there is complete segregation of the roots. Love resists the separation of unlikes and the clinging together of likes, by trying to keep unlike things mixed. The cosmos as we know it is

a result of intermediate phases between the two extremes of the triumph of one of the forces.[8]

Although Empedocles gives an account of the cosmos, cosmology is not his sole interest. Both fragments and testimonia show his keen attention to questions about perception and its role in knowledge, the workings of the body, and psychology. Like the Pythagoreans, Empedocles thought that how one lived was as important as one's theoretical commitments (and that the two were intimately connected). The ancient evidence seems to suggest that Empedocles was the author of two works, commonly called in modern scholarship the *Physics* and the *Purifications*, one cosmological and the other ethico-religious. The relation between the two works has been a matter of some controversy. In the 1990s new evidence from the Strasbourg Papyrus showed unequivocally that the cosmological and ethico-religious aspects of Empedocles' thought are inextricably intertwined (Martin and Primavesi 1999, Primavesi 2008, Kingsley 1995), although commentators still disagree about whether this new evidence supports the conclusion that there was a single poem combining both.[9] The correct philosophical understanding of the physical world and the correct way to live cannot be separated from one another in Empedocles' thought (a similar attitude appears in Heraclitus); one cannot fully understand the world without living correctly.[10] Like the Pythagorean, the Empedoclean way of life included dietary restrictions and a story of transmigrating *daimōns* who seem to have some kind of personal identity.

8. Presocratic Atomism

The pluralism of Anaxagoras and Empedocles maintained the Eleatic strictures on metaphysically acceptable basic entities (things that are and must be just what they are) by adopting an irreducible pluralism of stuffs meeting these standards that could pass on their qualities to items constructed from them. Ancient atomism responded more radically: what

is real is an infinite number of solid, uncuttable (*atomon*) units of matter. All atoms are made of the same stuff (solid matter, in itself otherwise indeterminate), differing from one another (according to Aristotle in Metaphysics 985b4-20=DK67A6) only in shape, position, arrangement. (Later sources say that atoms differ in weight; this is certainly true for post-Aristotelian atomism, but less likely for Presocratic atomism.) In addition, the Presocratic atomists, Leucippus and Democritus (Democritus was born in about 460 BCE in Abdera in Northern Greece, shortly after Socrates was born in Athens), enthusiastically endorsed the reality of the empty (or void).[11] The void is what separates atoms and allows for the differences noted above (except weight, which could not be accounted for by void, since void in an atom would make it divisible and, hence, not an atom) (Sedley 1982; see also Sedley 2008).

Like Anaxagoras, the atomists consider all phenomenal objects and characteristics as emerging from the background mixture; in the case of atomism, the mix of atoms and void (Wardy 1988). Everything is constructed of atoms and void: the shapes of the atoms and their arrangement with respect to each other (and the intervening void) give physical objects their apparent characteristics. As Democritus says: "By convention sweet and by convention bitter, by convention hot, by convention cold, by convention color: in reality atoms and void" (68B125 = B9). For example, Theophrastus says that the flavors differ according to the shapes of the atoms that compose various objects; thus "Democritus makes sweet that which is round and quite large, astringent that which large, rough, polygonal and not rounded" (*de Caus. Plant.* 6.1.6 = 68A129). Simplicius reports that things composed of sharp and very fine atoms in similar positions are hot and fiery; those composed of atoms with the opposite character come to be cold and watery (*in Phys.* 36.3–6 = 67A14). Moreover, Theophrastus reports that the atomists explain why iron is harder than lead but lighter; it is harder because of the uneven arrangements of the atoms that make it up, lighter because it contains

more void than lead. Lead, on the other hand, has less void than iron, but the even arrangement of the atoms makes lead easier to cut or to bend (*de Sens*. 61-63 = 68A135).

Adopting a strong distinction between appearance and reality, and denying the accuracy of appearances, as we see him do in the above quotation, Democritus was seen by some ancient sources (especially Sextus Empiricus) as a sort of skeptic, yet the evidence is unclear. It is true that Democritus is quoted as saying, "In truth we know nothing; for truth is in the depths" (68B117). So for him, the truth is not given in the appearances. Yet, even Sextus seems to agree that Democritus allows for knowledge:

> But in the *Rules* [Democritus] says that there are two kinds of knowing, one through the senses and the other through the understanding. The one through the understanding he calls genuine, witnessing to its trustworthiness in deciding truth; the one through the senses he names bastard, denying it steadfastness in the discernment of what is true. He says in these words, "There are two forms of knowing, one genuine and the other bastard. To the bastard belong all these: sight, hearing, smell, taste, touch. The other, the genuine, has been separated from this" [68B11]. Then preferring the genuine to the bastard, he continues, saying, "Whenever the bastard is no longer able to see more finely nor hear nor smell nor taste nor perceive by touch, but something finer…"

Thus Sextus suggests that the evidence of the senses, when properly interpreted by reason, can be taken as a guide to reality (the claim that "appearances are a sight of the unseen" is attributed to Democritus as well as to Anaxagoras). We just need to know how to follow this guide, through proper reasoning, so as to reach the truth—i.e., the theory of atoms and void (Lee 2005).

In addition to fragments advancing these metaphysical and physical doctrines, there are a number of ethical fragments attributed to Democritus (but the question of authenticity looms large here); although a passage reported in John Stobaeus seems to link moderation and cheerfulness with small measured movements in the soul and says that excess and deficiencies give rise to large movements (68B191), it is unclear whether or how these claims are related to the metaphysical aspects of atomism (Vlastos 1945 and 1946, Kahn 1985b). Democritus was identified in antiquity with the idea of "good cheer" (*euthumiē*) as the proper guiding objective in living one's life. In this, as in other aspects of his philosophy, he may have had some influence on the formation of Epicurus' philosophy a century later.

9. Diogenes of Apollonia and the Sophists

In the last part of the 5th century, Diogenes of Apollonia (active after 440 BCE) revived and revised the Milesian system of cosmology, claiming that "all the things that are are alterations from the same thing and are the same thing" (64B2); he identified this single basic substance with air, like Anaximenes more than a century before (Graham 2006, Laks 2008, 2008a). Diogenes takes care to give arguments for the reality and properties of his basic principle. In B2 he says that only things that are alike can affect one another. If there were a plurality of basic substances, each differing in what Diogenes calls their "own proper nature," there could be no interaction between them. Yet the evidence of the senses is clear: things mix and separate and interact with one another. Thus, all things must be forms of some one single thing. Like Anaxagoras, Diogenes claims that the cosmic system is ordered by intelligence, and he argues that that "which possesses intelligence (*noēsis*) is what human beings call air" (B5). Humans and animals live by breathing air, and are governed by it —in them air is both soul and intelligence, or mind (B4).

Moreover, Diogenes argues, air governs and rules all things and is god (B5). Thus, like Anaxagoras, Diogenes has a theory grounded in intelligence, although Diogenes is more fully committed to teleological explanations, insofar as he states explicitly that intelligence (*noēsis*) orders things in a good way (B3). In presenting his arguments, Diogenes fulfills his own requirement for a philosophical claim. In B1 he says, "In my opinion, anyone beginning a *logos* (account) ought to present a starting principle (*archē*) that is indisputable and a style that is simple and stately." He notes that his theory that air is soul and intelligence "will have been made clearly evident in this book" (B4).

Theophrastus says that Diogenes was the last of the physical philosophers, the *physiologoi*, or "inquirers into nature," as Aristotle called them; Diogenes Laertius gives that title to Archelaus, saying that he was the teacher of Socrates (*Lives* II.16-17). There was also another group of thinkers active about this time: the Sophists. Many of our views about this group have been shaped by Plato's aggressively negative assessment of them: in his dialogues Plato expressly contrasts the genuine philosopher, i.e., Socrates, with the Sophists, especially in their role as teachers of young men growing into their maturity (youths at the age when Socrates, too, engaged with them in his discussions). Modern scholarship (Woodruff and Gagarin 2008, Kerferd 1981, Guthrie 1969) has shown the diversity of their views. They were not completely uninterested in the theoretical problems that concerned others of the Presocratics. Gorgias of Leontini questioned the possibility of the certainty that Parmenides sought. In his *On Nature*, or *On what-is-not*, Gorgias claims that nothing satisfies Parmenides' requirements for what-is (Mansfeld 1985, Mourelatos 1987b, Palmer 1999, Caston 2002, Curd 2006). Protagoras, too, doubted the possibility of the strong theoretical knowledge that the Presocratics championed. The Sophists raised ethical and political questions: Does law or convention ground what is right, or is it a matter of nature? They traveled widely, sometimes serving as diplomats, and they were both

entertainers and teachers. They gave public displays of rhetoric (this contrasts with Diogenes of Apollonia's comments about his book, which seems to imply a more private enterprise)[12] and took on students, teaching both the art of rhetoric and the skills necessary for succeeding in Greek political life. With the Sophists, as with Socrates, interest in ethics and political thought becomes a more prominent aspect of Greek philosophy.

10. The Presocratic Legacy

The range of Presocratic thought shows that the first philosophers were not merely physicists (although they were certainly that). Their interests extended to religious and ethical thought, the nature of perception and understanding, mathematics, meteorology, the nature of explanation, and the roles of matter, form, causal mechanisms, and structure in the world. Almost all the Presocratics seemed to have something to say about embryology, and fragments of Diogenes and Empedocles show a keen interest in the structures of the body; the overlap between ancient philosophy and ancient medicine is of growing interest to scholars of early Greek thought (Longrigg 1963, van der Eijk 2008). Recent discoveries, such as the Derveni Papyrus (Betegh 2004, Kouremenos et al. 2006, Janko 2001, Laks and Most 1997), show that interest in and knowledge of the early philosophers was not necessarily limited to a small audience of rationalistic intellectuals. They passed on many of what later became the basic concerns of philosophy to Plato and Aristotle, and ultimately to the whole tradition of Western philosophical thought.

Bibliography

Primary Sources: Texts and Translations

Bollack, J., 1965 and 1969, *Empédocle*, vol. I, 1965; vols. II and III, 1969,

Plato

First published Sat Mar 20, 2004; substantive revision Tue Aug 1, 2017

Plato (429?–347 B.C.E.) is, by any reckoning, one of the most dazzling writers in the Western literary tradition and one of the most penetrating, wide-ranging, and influential authors in the history of philosophy. An Athenian citizen of high status, he displays in his works his absorption in the political events and intellectual movements of his time, but the questions he raises are so profound and the strategies he uses for tackling them so richly suggestive and provocative that educated readers of nearly every period have in some way been influenced by him, and in practically every age there have been philosophers who count themselves Platonists in some important respects. He was not the first thinker or writer to whom the word "philosopher" should be applied. But he was so self-conscious about how philosophy should be conceived, and what its scope and ambitions properly are, and he so transformed the intellectual currents with which he grappled, that the subject of philosophy, as it is often conceived—a rigorous and systematic examination of ethical, political, metaphysical, and epistemological issues, armed with a distinctive method—can be called his invention. Few other authors in the history of Western philosophy approximate him in depth and range: perhaps only Aristotle (who studied with him), Aquinas, and Kant would be generally agreed to be of the same rank.

- 1. Plato's central doctrines
- 2. Plato's puzzles
- 3. Dialogue, setting, character
- 4. Socrates
- 5. Plato's indirectness
- 6. Can we know Plato's mind?
- 7. Socrates as the dominant speaker

1. Plato's central doctrines

Many people associate Plato with a few central doctrines that are advocated in his writings: The world that appears to our senses is in some way defective and filled with error, but there is a more real and perfect realm, populated by entities (called "forms" or "ideas") that are eternal, changeless, and in some sense paradigmatic for the structure and character of the world presented to our senses. Among the most important of these abstract objects (as they are now called, because they are not located in space or time) are goodness, beauty, equality, bigness, likeness, unity, being, sameness, difference, change, and changelessness. (These terms —"goodness", "beauty", and so on—are often capitalized by those who write about Plato, in order to call attention to their exalted status; similarly for "Forms" and "Ideas.") The most fundamental distinction in Plato's philosophy is between the many observable objects that appear beautiful

(good, just, unified, equal, big) and the one object that is what beauty (goodness, justice, unity) really is, from which those many beautiful (good, just, unified, equal, big) things receive their names and their corresponding characteristics. Nearly every major work of Plato is, in some way, devoted to or dependent on this distinction. Many of them explore the ethical and practical consequences of conceiving of reality in this bifurcated way. We are urged to transform our values by taking to heart the greater reality of the forms and the defectiveness of the corporeal world. We must recognize that the soul is a different sort of object from the body—so much so that it does not depend on the existence of the body for its functioning, and can in fact grasp the nature of the forms far more easily when it is not encumbered by its attachment to anything corporeal. In a few of Plato's works, we are told that the soul always retains the ability to recollect what it once grasped of the forms, when it was disembodied prior to its possessor's birth (see especially *Meno*), and that the lives we lead are to some extent a punishment or reward for choices we made in a previous existence (see especially the final pages of *Republic*). But in many of Plato's writings, it is asserted or assumed that true philosophers—those who recognize how important it is to distinguish the one (the one thing that goodness is, or virtue is, or courage is) from the many (the many things that are called good or virtuous or courageous)— are in a position to become ethically superior to unenlightened human beings, because of the greater degree of insight they can acquire. To understand which things are good and why they are good (and if we are not interested in such questions, how can we become good?), we must investigate the form of good.

2. Plato's puzzles

Although these propositions are often identified by Plato's readers as forming a large part of the core of his philosophy, many of his greatest

admirers and most careful students point out that few, if any, of his writings can accurately be described as mere advocacy of a cut-and-dried group of propositions. Often Plato's works exhibit a certain degree of dissatisfaction and puzzlement with even those doctrines that are being recommended for our consideration. For example, the forms are sometimes described as hypotheses (see for example *Phaedo*). The form of good in particular is described as something of a mystery whose real nature is elusive and as yet unknown to anyone at all (*Republic*). Puzzles are raised—and not overtly answered—about how *any* of the forms can be known and how we are to talk about them without falling into contradiction (*Parmenides*), or about what it is to know anything (*Theaetetus*) or to name anything (*Cratylus*). When one compares Plato with some of the other philosophers who are often ranked with him— Aristotle, Aquinas, and Kant, for example—he can be recognized to be far more exploratory, incompletely systematic, elusive, and playful than they. That, along with his gifts as a writer and as a creator of vivid character and dramatic setting, is one of the reasons why he is often thought to be the ideal author from whom one should receive one's introduction to philosophy. His readers are not presented with an elaborate system of doctrines held to be so fully worked out that they are in no need of further exploration or development; instead, what we often receive from Plato is a few key ideas together with a series of suggestions and problems about how those ideas are to be interrogated and deployed. Readers of a Platonic dialogue are drawn into thinking for themselves about the issues raised, if they are to learn what the dialogue itself might be thought to say about them. Many of his works therefore give their readers a strong sense of philosophy as a living and unfinished subject (perhaps one that can never be completed) to which they themselves will have to contribute. All of Plato's works are in some way meant to leave further work for their readers, but among the ones that most conspicuously fall into this category

are: *Euthyphro*, *Laches*, *Charmides*, *Euthydemus*, *Theaetetus*, and *Parmenides*.

3. Dialogue, setting, character

There is another feature of Plato's writings that makes him distinctive among the great philosophers and colors our experience of him as an author. Nearly everything he wrote takes the form of a dialogue. (There is one striking exception: his *Apology*, which purports to be the speech that Socrates gave in his defense—the Greek word *apologia* means "defense"—when, in 399, he was legally charged and convicted of the crime of impiety. However, even there, Socrates is presented at one point addressing questions of a philosophical character to his accuser, Meletus, and responding to them. In addition, since antiquity, a collection of 13 letters has been included among his collected works, but their authenticity as compositions of Plato is not universally accepted among scholars, and many or most of them are almost certainly not his. Most of them purport to be the outcome of his involvement in the politics of Syracuse, a heavily populated Greek city located in Sicily and ruled by tyrants.)

We are of course familiar with the dialogue form through our acquaintance with the literary genre of drama. But Plato's dialogues do not try to create a fictional world for the purposes of telling a story, as many literary dramas do; nor do they invoke an earlier mythical realm, like the creations of the great Greek tragedians Aeschylus, Sophocles, and Euripides. Nor are they all presented in the form of a drama: in many of them, a single speaker narrates events in which he participated. They are philosophical discussions—"debates" would, in some cases, also be an appropriate word —among a small number of interlocutors, many of whom can be identified as real historical figures; and often they begin with a depiction of the setting of the discussion—a visit to a prison, a wealthy man's house, a celebration over drinks, a religious festival, a visit to the gymnasium, a

stroll outside the city's wall, a long walk on a hot day. As a group, they form vivid portraits of a social world, and are not purely intellectual exchanges between characterless and socially unmarked speakers. (At any rate, that is true of a large number of Plato's interlocutors. However, it must be added that in some of his works the speakers display little or no character. See, for example, *Sophist* and *Statesman*—dialogues in which a visitor from the town of Elea in Southern Italy leads the discussion; and *Laws*, a discussion between an unnamed Athenian and two named fictional characters, one from Crete and the other from Sparta.) In *many* of his dialogues (though not all), Plato is not only attempting to draw his readers into a discussion, but is also commenting on the social milieu that he is depicting, and criticizing the character and ways of life of his interlocutors. Some of the dialogues that most evidently fall into this category are *Protagoras*, *Gorgias*, *Hippias Major*, *Euthydemus*, and *Symposium*.

4. Socrates

There is one interlocutor who speaks in nearly all of Plato's dialogues, being completely absent only in *Laws*, which ancient testimony tells us was one of his latest works: that figure is Socrates. Like nearly everyone else who appears in Plato's works, he is not an invention of Plato: there really was a Socrates just as there really was a Crito, a Gorgias, a Thrasymachus, and a Laches. Plato was not the only author whose personal experience of Socrates led to the depiction of him as a character in one or more dramatic works. Socrates is one of the principal characters of Aristophanes' comedy, *Clouds*; and Xenophon, a historian and military leader, wrote, like Plato, both an *Apology* of Socrates (an account of Socrates' trial) and other works in which Socrates appears as a principal speaker. Furthermore, we have some fragmentary remains of dialogues written by other contemporaries of Socrates besides Plato and Xenophon

(Aeschines, Antisthenes, Eucleides, Phaedo), and these purport to describe conversations he conducted with others. So, when Plato wrote dialogues that feature Socrates as a principal speaker, he was both contributing to a genre that was inspired by the life of Socrates and participating in a lively literary debate about the kind of person Socrates was and the value of the intellectual conversations in which he was involved. Aristophanes' comic portrayal of Socrates is at the same time a bitter critique of him and other leading intellectual figures of the day (the 420s B.C.), but from Plato, Xenophon, and the other composers (in the 390's and later) of "Socratic discourses" (as Aristotle calls this body of writings) we receive a far more favorable impression.

Evidently, the historical Socrates was the sort of person who provoked in those who knew him, or knew of him, a profound response, and he inspired many of those who came under his influence to write about him. But the portraits composed by Aristophanes, Xenophon, and Plato are the ones that have survived intact, and they are therefore the ones that must play the greatest role in shaping our conception of what Socrates was like. Of these, *Clouds* has the least value as an indication of what was distinctive of Socrates' mode of philosophizing: after all, it is not intended as a philosophical work, and although it may contain a few lines that are characterizations of features unique to Socrates, for the most part it is an attack on a philosophical type—the long-haired, unwashed, amoral investigator into abstruse phenomena—rather than a depiction of Socrates himself. Xenophon's depiction of Socrates, whatever its value as historical testimony (which may be considerable), is generally thought to lack the philosophical subtlety and depth of Plato's. At any rate, no one (certainly not Xenophon himself) takes Xenophon to be a major philosopher in his own right; when we read his Socratic works, we are not encountering a great philosophical mind. But that is what we experience when we read Plato. We may read Plato's Socratic dialogues because we are (as Plato evidently wanted us to be) interested in who Socrates was and what he

stood for, but even if we have little or no desire to learn about the historical Socrates, we will want to read Plato because in doing so we are encountering an author of the greatest philosophical significance. No doubt he in some way borrowed in important ways from Socrates, though it is not easy to say where to draw the line between him and his teacher (more about this below in section 12). But it is widely agreed among scholars that Plato is not a mere transcriber of the words of Socrates (any more than Xenophon or the other authors of Socratic discourses). His use of a figure called "Socrates" in so many of his dialogues should not be taken to mean that Plato is merely preserving for a reading public the lessons he learned from his teacher.

5. Plato's indirectness

Socrates, it should be kept in mind, does not appear in all of Plato's works. He makes no appearance in *Laws*, and there are several dialogues (*Sophist*, *Statesman*, *Timaeus*) in which his role is small and peripheral, while some other figure dominates the conversation or even, as in the *Timaeus* and *Critias*, presents a long and elaborate, continuous discourse of their own. Plato's dialogues are not a static literary form; not only do his topics vary, not only do his speakers vary, but the role played by questions and answers is never the same from one dialogue to another. (*Symposium*, for example, is a series of speeches, and there are also lengthy speeches in *Apology*, *Menexenus*, *Protagoras*, *Crito*, *Phaedrus*, *Timaeus*, and *Critias*; in fact, one might reasonably question whether these works are properly called dialogues). But even though Plato constantly adapted "the dialogue form" (a commonly used term, and convenient enough, so long as we do not think of it as an unvarying unity) to suit his purposes, it is striking that throughout his career as a writer he never engaged in a form of composition that was widely used in his time and was soon to become the standard mode of philosophical address: Plato never became a writer of

philosophical treatises, even though the writing of treatises (for example, on rhetoric, medicine, and geometry) was a common practice among his predecessors and contemporaries. (The closest we come to an exception to this generalization is the seventh letter, which contains a brief section in which the author, Plato or someone pretending to be him, commits himself to several philosophical points—while insisting, at the same time, that no philosopher will write about the deepest matters, but will communicate his thoughts only in private discussion with selected individuals. As noted above, the authenticity of Plato's letters is a matter of great controversy; and in any case, the author of the seventh letter declares his opposition to the writing of philosophical books. Whether Plato wrote it or not, it cannot be regarded as a philosophical treatise, and its author did not wish it to be so regarded.) In all of his writings—except in the letters, if any of them are genuine—Plato never speaks to his audience directly and in his own voice. Strictly speaking, he does not himself affirm anything in his dialogues; rather, it is the interlocutors in his dialogues who are made by Plato to do all of the affirming, doubting, questioning, arguing, and so on. Whatever he wishes to communicate to us is conveyed indirectly.

6. Can we know Plato's mind?

This feature of Plato's works raises important questions about how they are to be read, and has led to considerable controversy among those who study his writings. Since he does not himself affirm anything in any of his dialogues, can we ever be on secure ground in attributing a philosophical doctrine to him (as opposed to one of his characters)? Did he himself have philosophical convictions, and can we discover what they were? Are we justified in speaking of "the philosophy of Plato"? Or, if we attribute some view to Plato himself, are we being unfaithful to the spirit in which he intended the dialogues to be read? Is his whole point, in refraining from writing treatises, to discourage the readers of his works from asking what

their author believes and to encourage them instead simply to consider the plausibility or implausibility of what his characters are saying? Is that why Plato wrote dialogues? If not for this reason, then what *was* his purpose in refraining from addressing his audience in a more direct way? There are other important questions about the particular shape his dialogues take: for example, why does Socrates play such a prominent role in so many of them, and why, in some of these works, does Socrates play a smaller role, or none at all?

Once these questions are raised and their difficulty acknowledged, it is tempting, in reading Plato's works and reflecting upon them, to adopt a strategy of extreme caution. Rather than commit oneself to any hypothesis about what he is trying to communicate to his readers, one might adopt a stance of neutrality about his intentions, and confine oneself to talking only about what is said by his *dramatis personae*. One cannot be faulted, for example, if one notes that, in Plato's *Republic*, Socrates argues that justice in the soul consists in each part of the soul doing its own. It is equally correct to point out that other principal speakers in that work, Glaucon and Adeimantus, accept the arguments that Socrates gives for that definition of justice. Perhaps there is no need for us to say more—to say, for example, that Plato himself agrees that this is how justice should be defined, or that Plato himself accepts the arguments that Socrates gives in support of this definition. And we might adopt this same "minimalist" approach to *all* of Plato's works. After all, is it of any importance to discover what went on inside his head as he wrote—to find out whether he himself endorsed the ideas he put in the mouths of his characters, whether they constitute "the philosophy of Plato"? Should we not read his works for their intrinsic philosophical value, and not as tools to be used for entering into the mind of their author? We know what Plato's characters say—and isn't that all that we need, for the purpose of engaging with his works philosophically?

But the fact that we know what Plato's characters *say* does not show that by refusing to entertain any hypotheses about what the author of these works is trying to communicate to his readers we can understand what those characters *mean* by what they say. We should not lose sight of this obvious fact: it is Plato, not any of his *dramatis personae*, who is reaching out to a readership and trying to influence their beliefs and actions by means of his literary actions. When we ask whether an argument put forward by a character in Plato's works should be read as an effort to persuade us of its conclusion, or is better read as a revelation of how foolish that speaker is, we are asking about what *Plato* as author (not that character) is trying to lead us to believe, through the writing that he is presenting to our attention. We need to interpret the work itself to find out what it, or Plato the author, is saying. Similarly, when we ask how a word that has several different senses is best understood, we are asking what Plato means to communicate to us through the speaker who uses that word. We should not suppose that we can derive much philosophical value from Plato's writings if we refuse to entertain any thoughts about what use he intends us to make of the things his speakers say. Penetrating the mind of Plato and comprehending what his interlocutors mean by what they say are not two separate tasks but one, and if we do not ask what his interlocutors mean by what they say, and what the dialogue itself indicates we should think about what they mean, we will not profit from reading his dialogues.

Furthermore, the dialogues have certain characteristics that are most easily explained by supposing that Plato is using them as vehicles for inducing his readers to become convinced (or more convinced than they already are) of certain propositions—for example, that there are forms, that the soul is not corporeal, that knowledge can be acquired only by means of a study of the forms, and so on. Why, after all, did Plato write so many works (for example: *Phaedo*, *Symposium*, *Republic*, *Phaedrus*, *Theaetetus*, *Sophist*, *Statesman*, *Timaeus*, *Philebus*, *Laws*) in which one character

dominates the conversation (often, but not always, Socrates) and convinces the other speakers (at times, after encountering initial resistance) that they should accept or reject certain conclusions, on the basis of the arguments presented? The only plausible way of answering that question is to say that these dialogues were intended by Plato to be devices by which he might induce the audience for which they are intended to reflect on and accept the arguments and conclusions offered by his principal interlocutor. (It is noteworthy that in *Laws*, the principal speaker—an unnamed visitor from Athens—proposes that laws should be accompanied by "preludes" in which their philosophical basis is given as full an explanation as possible. The educative value of written texts is thus explicitly acknowledged by Plato's dominant speaker. If preludes can educate a whole citizenry that is prepared to learn from them, then surely Plato thinks that other sorts of written texts—for example, his own dialogues—can also serve an educative function.)

This does not mean that Plato thinks that his readers can become wise simply by reading and studying his works. On the contrary, it is highly likely that he wanted all of his writings to be supplementary aids to philosophical conversation: in one of his works, he has Socrates warn his readers against relying solely on books, or taking them to be authoritative. They are, Socrates says, best used as devices that stimulate the readers' memory of discussions they have had (*Phaedrus* 274e-276d). In those face-to-face conversations with a knowledgeable leader, positions are taken, arguments are given, and conclusions are drawn. Plato's writings, he implies in this passage from *Phaedrus*, will work best when conversational seeds have already been sown for the arguments they contain.

7. Socrates as the dominant speaker

If we take Plato to be trying to persuade us, in many of his works, to accept the conclusions arrived at by his principal interlocutors (or to persuade us of the refutations of their opponents), we can easily explain why he so often chooses Socrates as the dominant speaker in his dialogues. Presumably the contemporary audience for whom Plato was writing included many of Socrates' admirers. They would be predisposed to think that a character called "Socrates" would have all of the intellectual brilliance and moral passion of the historical person after whom he is named (especially since Plato often makes special efforts to give his "Socrates" a life-like reality, and has him refer to his trial or to the characteristics by which he was best known); and the aura surrounding the character called "Socrates" would give the words he speaks in the dialogue considerable persuasive power. Furthermore, if Plato felt strongly indebted to Socrates for many of his philosophical techniques and ideas, that would give him further reason for assigning a dominant role to him in many of his works. (More about this in section 12.)

Of course, there are other more speculative possible ways of explaining why Plato so often makes Socrates his principal speaker. For example, we could say that Plato was trying to undermine the reputation of the historical Socrates by writing a series of works in which a figure called "Socrates" manages to persuade a group of naïve and sycophantic interlocutors to accept absurd conclusions on the basis of sophistries. But anyone who has read some of Plato's works will quickly recognize the utter implausibility of that alternative way of reading them. Plato could have written into his works clear signals to the reader that the arguments of Socrates do not work, and that his interlocutors are foolish to accept them. But there are many signs in such works as *Meno*, *Phaedo*, *Republic*, and *Phaedrus* that point in the opposite direction. (And the great admiration Plato feels for Socrates is also evident from his *Apology*.) The reader is given every encouragement to believe that the reason why Socrates is successful in persuading his interlocutors (on those occasions

when he does succeed) is that his arguments are powerful ones. The reader, in other words, is being encouraged by the author to accept those arguments, if not as definitive then at least as highly arresting and deserving of careful and full positive consideration. When we interpret the dialogues in this way, we cannot escape the fact that we are entering into the mind of Plato, and attributing to him, their author, a positive evaluation of the arguments that his speakers present to each other.

8. Links between the dialogues

There is a further reason for entertaining hypotheses about what Plato intended and believed, and not merely confining ourselves to observations about what sorts of people his characters are and what they say to each other. When we undertake a serious study of Plato, and go beyond reading just one of his works, we are inevitably confronted with the question of how we are to link the work we are currently reading with the many others that Plato composed. Admittedly, many of his dialogues make a fresh start in their setting and their interlocutors: typically, Socrates encounters a group of people many of whom do not appear in any other work of Plato, and so, as an author, he needs to give his readers some indication of their character and social circumstances. But often Plato's characters make statements that would be difficult for readers to understand unless they had already read one or more of his other works. For example, in *Phaedo* (73a-b), Socrates says that one argument for the immortality of the soul derives from the fact that when people are asked certain kinds of questions, and are aided with diagrams, they answer in a way that shows that they are not learning afresh from the diagrams or from information provided in the questions, but are drawing their knowledge of the answers from within themselves. That remark would be of little worth for an audience that had not already read *Meno*. Several pages later, Socrates tells his interlocutors that his argument about our prior knowledge of equality itself (the form of

equality) applies no less to other forms—to the beautiful, good, just, pious and to all the other things that are involved in their asking and answering of questions (75d). This reference to asking and answering questions would not be well understood by a reader who had not yet encountered a series of dialogues in which Socrates asks his interlocutors questions of the form, "What is X?" (*Euthyphro*: what is piety? *Laches*: what is courage? *Charmides*: What is moderation? *Hippias Major*: what is beauty?). Evidently, Plato is assuming that readers of *Phaedo* have already read several of his other works, and will bring to bear on the current argument all of the lessons that they have learned from them. In some of his writings, Plato's characters refer ahead to the continuation of their conversations on another day, or refer back to conversations they had recently: thus Plato signals to us that we should read *Theaetetus*, *Sophist*, and *Statesman* sequentially; and similarly, since the opening of *Timaeus* refers us back to *Republic*, Plato is indicating to his readers that they must seek some connection between these two works.

These features of the dialogues show Plato's awareness that he cannot entirely start from scratch in every work that he writes. He will introduce new ideas and raise fresh difficulties, but he will also expect his readers to have already familiarized themselves with the conversations held by the interlocutors of other dialogues—even when there is some alteration among those interlocutors. (Meno does not re-appear in *Phaedo*; Timaeus was not among the interlocutors of *Republic*.) Why does Plato have his dominant characters (Socrates, the Eleatic visitor) reaffirm some of the same points from one dialogue to another, and build on ideas that were made in earlier works? If the dialogues were merely meant as provocations to thought—mere exercises for the mind—there would be no need for Plato to identify his leading characters with a consistent and ever-developing doctrine. For example, Socrates continues to maintain, over a large number of dialogues, that there are such things as forms—and there is no better explanation for this continuity than to suppose that Plato is

recommending that doctrine to his readers. Furthermore, when Socrates is replaced as the principal investigator by the visitor from Elea (in *Sophist* and *Statesman*), the existence of forms continues to be taken for granted, and the visitor criticizes any conception of reality that excludes such incorporeal objects as souls and forms. The Eleatic visitor, in other words, upholds a metaphysics that is, in many respects, like the one that Socrates is made to defend. Again, the best explanation for this continuity is that Plato is using both characters—Socrates and the Eleatic visitor—as devices for the presentation and defense of a doctrine that he embraces and wants his readers to embrace as well.

9. Does Plato change his mind about forms?

This way of reading Plato's dialogues does not presuppose that he never changes his mind about anything—that whatever any of his main interlocutors uphold in one dialogue will continue to be presupposed or affirmed elsewhere without alteration. It is, in fact, a difficult and delicate matter to determine, on the basis of our reading of the dialogues, whether Plato means to modify or reject in one dialogue what he has his main interlocutor affirm in some other. One of the most intriguing and controversial questions about his treatment of the forms, for example, is whether he concedes that his conception of those abstract entities is vulnerable to criticism; and, if so, whether he revises some of the assumptions he had been making about them, or develops a more elaborate picture of them that allows him to respond to that criticism. In *Parmenides*, the principal interlocutor (not Socrates—he is here portrayed as a promising, young philosopher in need of further training—but rather the pre-Socratic from Elea who gives the dialogue its name: Parmenides) subjects the forms to withering criticism, and then consents to conduct an inquiry into the nature of oneness that has no overt connection to his critique of the forms. Does the discussion of oneness (a baffling series of

contradictions—or at any rate, propositions that seem, on the surface, to be contradictions) in some way help address the problems raised about forms? That is one way of reading the dialogue. And if we do read it in this way, does that show that Plato has changed his mind about some of the ideas about forms he inserted into earlier dialogues? Can we find dialogues in which we encounter a "new theory of forms"—that is, a way of thinking of forms that carefully steers clear of the assumptions about forms that led to Parmenides' critique? It is not easy to say. But we cannot even raise this as an issue worth pondering unless we presuppose that behind the dialogues there stands a single mind that is using these writings as a way of hitting upon the truth, and of bringing that truth to the attention of others. If we find Timaeus (the principal interlocutor of the dialogue named after him) and the Eleatic visitor of the *Sophist* and *Statesman* talking about forms in a way that is entirely consistent with the way Socrates talks about forms in *Phaedo* and *Republic*, then there is only one reasonable explanation for that consistency: Plato believes that their way of talking about forms is correct, or is at least strongly supported by powerful considerations. If, on the other hand, we find that Timaeus or the Eleatic visitor talks about forms in a way that does not harmonize with the way Socrates conceives of those abstract objects, in the dialogues that assign him a central role as director of the conversation, then the most plausible explanation for these discrepancies is that Plato has changed his mind about the nature of these entities. It would be implausible to suppose that Plato himself had no convictions about forms, and merely wants to give his readers mental exercise by composing dialogues in which different leading characters talk about these objects in discordant ways.

10. Does Plato change his mind about politics?

The same point—that we must view the dialogues as the product of a single mind, a single philosopher, though perhaps one who changes his

mind—can be made in connection with the politics of Plato's works.

It is noteworthy, to begin with, that Plato is, among other things, a *political* philosopher. For he gives expression, in several of his writings (particular *Phaedo*), to a yearning to escape from the tawdriness of ordinary human relations. (Similarly, he evinces a sense of the ugliness of the sensible world, whose beauty pales in comparison with that of the forms.) Because of this, it would have been all too easy for Plato to turn his back entirely on practical reality, and to confine his speculations to theoretical questions. Some of his works—*Parmenides* is a stellar example —do confine themselves to exploring questions that seem to have no bearing whatsoever on practical life. But it is remarkable how few of his works fall into this category. Even the highly abstract questions raised in *Sophist* about the nature of being and not-being are, after all, embedded in a search for the definition of sophistry; and thus they call to mind the question whether Socrates should be classified as a sophist—whether, in other words, sophists are to be despised and avoided. In any case, despite the great sympathy Plato expresses for the desire to shed one's body and live in an incorporeal world, he devotes an enormous amount of energy to the task of understanding the world we live in, appreciating its limited beauty, and improving it.

His tribute to the mixed beauty of the sensible world, in *Timaeus*, consists in his depiction of it as the outcome of divine efforts to mold reality in the image of the forms, using simple geometrical patterns and harmonious arithmetic relations as building blocks. The desire to transform human relations is given expression in a far larger number of works. Socrates presents himself, in Plato's *Apology*, as a man who does not have his head in the clouds (that is part of Aristophanes' charge against him in *Clouds*). He does not want to escape from the everyday world but to make it better. He presents himself, in *Gorgias*, as the only Athenian who has tried his hand at the true art of politics.

Similarly, the Socrates of *Republic* devotes a considerable part of his discussion to the critique of ordinary social institutions—the family, private property, and rule by the many. The motivation that lies behind the writing of this dialogue is the desire to transform (or, at any rate, to improve) political life, not to escape from it (although it is acknowledged that the desire to escape is an honorable one: the best sort of rulers greatly prefer the contemplation of divine reality to the governance of the city). And if we have any further doubts that Plato does take an interest in the practical realm, we need only turn to *Laws*. A work of such great detail and length about voting procedures, punishments, education, legislation, and the oversight of public officials can only have been produced by someone who wants to contribute something to the improvement of the lives we lead in this sensible and imperfect realm. Further evidence of Plato's interest in practical matters can be drawn from his letters, if they are genuine. In most of them, he presents himself as having a deep interest in educating (with the help of his friend, Dion) the ruler of Syracuse, Dionysius II, and thus reforming that city's politics.

Just as any attempt to understand Plato's views about forms must confront the question whether his thoughts about them developed or altered over time, so too our reading of him as a political philosopher must be shaped by a willingness to consider the possibility that he changed his mind. For example, on any plausible reading of *Republic*, Plato evinces a deep antipathy to rule by the many. Socrates tells his interlocutors that the only politics that should engage them are those of the anti-democratic regime he depicts as the paradigm of a good constitution. And yet in *Laws*, the Athenian visitor proposes a detailed legislative framework for a city in which non-philosophers (people who have never heard of the forms, and have not been trained to understand them) are given considerable powers as rulers. Plato would not have invested so much time in the creation of this comprehensive and lengthy work, had he not believed that the creation of a political community ruled by those who are philosophically

unenlightened is a project that deserves the support of his readers. Has Plato changed his mind, then? Has he re-evaluated the highly negative opinion he once held of those who are innocent of philosophy? Did he at first think that the reform of existing Greek cities, with all of their imperfections, is a waste of time—but then decide that it is an endeavor of great value? (And if so, what led him to change his mind?) Answers to these questions can be justified only by careful attention to what he has his interlocutors say. But it would be utterly implausible to suppose that these developmental questions need not be raised, on the grounds that *Republic* and *Laws* each has its own cast of characters, and that the two works therefore cannot come into contradiction with each other. According to this hypothesis (one that must be rejected), because it is Socrates (not Plato) who is critical of democracy in *Republic*, and because it is the Athenian visitor (not Plato) who recognizes the merits of rule by the many in *Laws*, there is no possibility that the two dialogues are in tension with each other. Against this hypothesis, we should say: Since both *Republic* and *Laws* are works in which Plato is trying to move his readers towards certain conclusions, by having them reflect on certain arguments—these dialogues are not barred from having this feature by their use of interlocutors—it would be an evasion of our responsibility as readers and students of Plato not to ask whether what one of them advocates is compatible with what the other advocates. If we answer that question negatively, we have some explaining to do: what led to this change? Alternatively, if we conclude that the two works are compatible, we must say why the appearance of conflict is illusory.

11. The historical Socrates: early, middle, and late dialogues

Many contemporary scholars find it plausible that when Plato embarked on his career as a philosophical writer, he composed, in addition to his

Apology of Socrates, a number of short ethical dialogues that contain little or nothing in the way of positive philosophical doctrine, but are mainly devoted to portraying the way in which Socrates punctured the pretensions of his interlocutors and forced them to realize that they are unable to offer satisfactory definitions of the ethical terms they used, or satisfactory arguments for their moral beliefs. According to this way of placing the dialogues into a rough chronological order—associated especially with Gregory Vlastos's name (see especially his *Socrates Ironist and Moral Philosopher*, chapters 2 and 3)—Plato, at this point of his career, was content to use his writings primarily for the purpose of preserving the memory of Socrates and making plain the superiority of his hero, in intellectual skill and moral seriousness, to all of his contemporaries—particularly those among them who claimed to be experts on religious, political, or moral matters. Into this category of early dialogues (they are also sometimes called "Socratic" dialogues, possibly without any intended chronological connotation) are placed: *Charmides*, *Crito*, *Euthydemus*, *Euthyphro*, *Gorgias*, *Hippias Major*, *Hippias Minor*, *Ion*, *Laches*, *Lysis*, and *Protagoras*, (Some scholars hold that we can tell which of these come later during Plato's early period. For example, it is sometimes said that *Protagoras* and *Gorgias* are later, because of their greater length and philosophical complexity. Other dialogues—for example, *Charmides* and *Lysis*—are thought not to be among Plato's earliest within this early group, because in them Socrates appears to be playing a more active role in shaping the progress of the dialogue: that is, he has more ideas of his own.) In comparison with many of Plato's other dialogues, these "Socratic" works contain little in the way of metaphysical, epistemological, or methodological speculation, and they therefore fit well with the way Socrates characterizes himself in Plato's *Apology*: as a man who leaves investigations of high falutin' matters (which are "in the sky and below the earth") to wiser heads, and confines all of his investigations to the question how one should live one's life. Aristotle describes Socrates

as someone whose interests were restricted to only one branch of philosophy—the realm of the ethical; and he also says that he was in the habit of asking definitional questions to which he himself lacked answers (*Metaphysics* 987b1, *Sophistical Refutations* 183b7). That testimony gives added weight to the widely accepted hypothesis that there is a group of dialogues—the ones mentioned above as his early works, whether or not they were all written early in Plato's writing career—in which Plato used the dialogue form as a way of portraying the philosophical activities of the historical Socrates (although, of course, he might also have used them in other ways as well—for example to suggest and begin to explore philosophical difficulties raised by them).

But at a certain point—so says this hypothesis about the chronology of the dialogues—Plato began to use his works to advance ideas that were his own creations rather than those of Socrates, although he continued to use the name "Socrates" for the interlocutor who presented and argued for these new ideas. The speaker called "Socrates" now begins to move beyond and depart from the historical Socrates: he has views about the methodology that should be used by philosophers (a methodology borrowed from mathematics), and he argues for the immortality of the soul and the existence and importance of the forms of beauty, justice, goodness, and the like. (By contrast, in *Apology* Socrates says that no one knows what becomes of us after we die.) *Phaedo* is often said to be the dialogue in which Plato first comes into his own as a philosopher who is moving far beyond the ideas of his teacher (though it is also commonly said that we see a new methodological sophistication and a greater interest in mathematical knowledge in *Meno*). Having completed all of the dialogues that, according to this hypothesis, we characterize as early, Plato widened the range of topics to be explored in his writings (no longer confining himself to ethics), and placed the theory of forms (and related ideas about language, knowledge, and love) at the center of his thinking. In these works of his "middle" period—for example, in *Phaedo*, *Cratylus*,

Symposium, *Republic*, and *Phaedrus*—there is both a change of emphasis and of doctrine. The focus is no longer on ridding ourselves of false ideas and self-deceit; rather, we are asked to accept (however tentatively) a radical new conception of ourselves (now divided into three parts), our world—or rather, our two worlds—and our need to negotiate between them. Definitions of the most important virtue terms are finally proposed in *Republic* (the search for them in some of the early dialogues having been unsuccessful): Book I of this dialogue is a portrait of how the historical Socrates might have handled the search for a definition of justice, and the rest of the dialogue shows how the new ideas and tools discovered by Plato can complete the project that his teacher was unable to finish. Plato continues to use a figure called "Socrates" as his principal interlocutor, and in this way he creates a sense of continuity between the methods, insights, and ideals of the historical Socrates and the new Socrates who has now become a vehicle for the articulation of his own new philosophical outlook. In doing so, he acknowledges his intellectual debt to his teacher and appropriates for his own purposes the extraordinary prestige of the man who was the wisest of his time.

This hypothesis about the chronology of Plato's writings has a third component: it does not place his works into either of only two categories —the early or "Socratic" dialogues, and all the rest—but works instead with a threefold division of early, middle, and late. That is because, following ancient testimony, it has become a widely accepted assumption that *Laws* is one of Plato's last works, and further that this dialogue shares a great many stylistic affinities with a small group of others: *Sophist*, *Statesman*, *Timaeus*, *Critias*, and *Philebus*. These five dialogues together with *Laws* are generally agreed to be his late works, because they have much more in common with each other, when one counts certain stylistic features apparent only to readers of Plato's Greek, than with any of Plato's other works. (Computer counts have aided these stylometric studies, but

the isolation of a group of six dialogues by means of their stylistic commonalities was recognized in the nineteenth century.)

It is not at all clear whether there are one or more *philosophical* affinities among this group of six dialogues—that is, whether the philosophy they contain is sharply different from that of all of the other dialogues. Plato does nothing to encourage the reader to view these works as a distinctive and separate component of his thinking. On the contrary, he links *Sophist* with *Theaetetus* (the conversations they present have a largely overlapping cast of characters, and take place on successive days) no less than *Sophist* and *Statesman*. *Sophist* contains, in its opening pages, a reference to the conversation of *Parmenides*—and perhaps Plato is thus signaling to his readers that they should bring to bear on *Sophist* the lessons that are to be drawn from *Parmenides*. Similarly, *Timaeus* opens with a reminder of some of the principal ethical and political doctrines of *Republic*. It could be argued, of course, that when one looks beyond these stage-setting devices, one finds significant philosophical changes in the six late dialogues, setting this group off from all that preceded them. But there is no consensus that they should be read in this way. Resolving this issue requires intensive study of the content of Plato's works. So, although it is widely accepted that the six dialogues mentioned above belong to Plato's latest period, there is, as yet, no agreement among students of Plato that these six form a distinctive stage in his philosophical development.

In fact, it remains a matter of dispute whether the division of Plato's works into three periods—early, middle, late—does correctly indicate the order of composition, and whether it is a useful tool for the understanding of his thought (See Cooper 1997, vii–xxvii). Of course, it would be wildly implausible to suppose that Plato's writing career began with such complex works as *Laws*, *Parmenides*, *Phaedrus*, or *Republic*. In light of widely accepted assumptions about how most philosophical minds develop, it is likely that when Plato started writing philosophical works

some of the shorter and simpler dialogues were the ones he composed: *Laches*, or *Crito*, or *Ion* (for example). (Similarly, *Apology* does not advance a complex philosophical agenda or presuppose an earlier body of work; so that too is likely to have been composed near the beginning of Plato's writing career.) Even so, there is no good reason to eliminate the hypothesis that throughout much of his life Plato devoted himself to writing two sorts of dialogues at the same time, moving back and forth between them as he aged: on the one hand, introductory works whose primary purpose is to show readers the difficulty of apparently simple philosophical problems, and thereby to rid them of their pretensions and false beliefs; and on the other hand, works filled with more substantive philosophical theories supported by elaborate argumentation. Moreover, one could point to features of many of the "Socratic" dialogues that would justify putting them in the latter category, even though the argumentation does not concern metaphysics or methodology or invoke mathematics —*Gorgias*, *Protagoras*, *Lysis*, *Euthydemus*, *Hippias Major* among them.

Plato makes it clear that both of these processes, one preceding the other, must be part of one's philosophical education. One of his deepest methodological convictions (affirmed in *Meno*, *Theaetetus*, and *Sophist*) is that in order to make intellectual progress we must recognize that knowledge cannot be acquired by passively receiving it from others: rather, we must work our way through problems and assess the merits of competing theories with an independent mind. Accordingly, some of his dialogues are primarily devices for breaking down the reader's complacency, and that is why it is essential that they come to no positive conclusions; others are contributions to theory-construction, and are therefore best absorbed by those who have already passed through the first stage of philosophical development. We should not assume that Plato could have written the preparatory dialogues only at the earliest stage of his career. Although he may well have begun his writing career by taking up that sort of project, he may have continued writing these "negative"

works at later stages, at the same time that he was composing his theory-constructing dialogues. For example although both *Euthydemus* and *Charmides* are widely assumed to be early dialogues, they might have been written around the same time as *Symposium* and *Republic*, which are generally assumed to be compositions of his middle period—or even later.

No doubt, some of the works widely considered to be early really are such. But it is an open question which and how many of them are. At any rate, it is clear that Plato continued to write in a "Socratic" and "negative" vein even after he was well beyond the earliest stages of his career: *Theaetetus* features a Socrates who is even more insistent upon his ignorance than are the dramatic representations of Socrates in briefer and philosophically less complex works that are reasonably assumed to be early; and like many of those early works, *Theaetetus* seeks but does not find the answer to the "what is it?" question that it relentlessly pursues—"What is knowledge?" Similarly, *Parmenides*, though certainly not an early dialogue, is a work whose principal aim is to puzzle the reader by the presentation of arguments for apparently contradictory conclusions; since it does not tell us how it is possible to accept all of those conclusions, its principal effect on the reader is similar to that of dialogues (many of them no doubt early) that reach only negative conclusions. Plato uses this educational device— provoking the reader through the presentation of opposed arguments, and leaving the contradiction unresolved—in *Protagoras* (often considered an early dialogue) as well. So it is clear that even after he was well beyond the earliest stages of his thinking, he continued to assign himself the project of writing works whose principal aim is the presentation of unresolved difficulties. (And, just as we should recognize that puzzling the reader continues to be his aim even in later works, so too we should not overlook the fact that there is some substantive theory-construction in the ethical works that are simple enough to have been early compositions: *Ion*, for example, affirms a theory of poetic inspiration; and *Crito* sets out the

conditions under which a citizen acquires an obligation to obey civic commands. Neither ends in failure.)

If we are justified in taking Socrates' speech in Plato's *Apology* to constitute reliable evidence about what the historical Socrates was like, then whatever we find in Plato's other works that is of a piece with that speech can also be safely attributed to Socrates. So understood, Socrates was a moralist but (unlike Plato) not a metaphysician or epistemologist or cosmologist. That fits with Aristotle's testimony, and Plato's way of choosing the dominant speaker of his dialogues gives further support to this way of distinguishing between him and Socrates. The number of dialogues that are dominated by a Socrates who is spinning out elaborate philosophical doctrines is remarkably small: *Phaedo*, *Republic*, *Phaedrus*, and *Philebus*. All of them are dominated by ethical issues: whether to fear death, whether to be just, whom to love, the place of pleasure. Evidently, Plato thinks that it is appropriate to make Socrates the major speaker in a dialogue that is filled with positive content only when the topics explored in that work primarily have to do with the ethical life of the individual. (The political aspects of *Republic* are explicitly said to serve the larger question whether any individual, no matter what his circumstances, should be just.) When the doctrines he wishes to present systematically become primarily metaphysical, he turns to a visitor from Elea (*Sophist*, *Statesman*); when they become cosmological, he turns to Timaeus; when they become constitutional, he turns, in *Laws*, to a visitor from Athens (and he then eliminates Socrates entirely). In effect, Plato is showing us: although he owes a great deal to the ethical insights of Socrates, as well as to his method of puncturing the intellectual pretensions of his interlocutors by leading them into contradiction, he thinks he should not put into the mouth of his teacher too elaborate an exploration of ontological, or cosmological, or political themes, because Socrates refrained from entering these domains. This may be part of the explanation why he has Socrates put into the mouth of the personified Laws of Athens the theory

advanced in *Crito*, which reaches the conclusion that it would be unjust for him to escape from prison. Perhaps Plato is indicating, at the point where these speakers enter the dialogue, that none of what is said here is in any way derived from or inspired by the conversation of Socrates.

Just as we should reject the idea that Plato must have made a decision, at a fairly early point in his career, no longer to write one kind of dialogue (negative, destructive, preparatory) and to write only works of elaborate theory-construction; so we should also question whether he went through an early stage during which he refrained from introducing into his works any of his own ideas (if he had any), but was content to play the role of a faithful portraitist, representing to his readers the life and thought of Socrates. It is unrealistic to suppose that someone as original and creative as Plato, who probably began to write dialogues somewhere in his thirties (he was around 28 when Socrates was killed), would have started his compositions with no ideas of his own, or, having such ideas, would have decided to suppress them, for some period of time, allowing himself to think for himself only later. (What would have led to such a decision?) We should instead treat the moves made in the dialogues, even those that are likely to be early, as Platonic inventions—derived, no doubt, by Plato's reflections on and transformations of the key themes of Socrates that he attributes to Socrates in *Apology*. That speech indicates, for example, that the kind of religiosity exhibited by Socrates was unorthodox and likely to give offense or lead to misunderstanding. It would be implausible to suppose that Plato simply concocted the idea that Socrates followed a divine sign, especially because Xenophon too attributes this to his Socrates. But what of the various philosophical moves rehearsed in *Euthyphro*—the dialogue in which Socrates searches, unsuccessfully, for an understanding of what piety is? We have no good reason to think that in writing this work Plato adopted the role of a mere recording device, or something close to it (changing a word here and there, but for the most part simply recalling what he heard Socrates say, as he made his way to

court). It is more likely that Plato, having been inspired by the unorthodoxy of Socrates' conception of piety, developed, on his own, a series of questions and answers designed to show his readers how difficult it is to reach an understanding of the central concept that Socrates' fellow citizens relied upon when they condemned him to death. The idea that it is important to search for definitions may have been Socratic in origin. (After all, Aristotle attributes this much to Socrates.) But the twists and turns of the arguments in *Euthyphro* and other dialogues that search for definitions are more likely to be the products of Plato's mind than the content of any conversations that really took place.

12. Why dialogues?

It is equally unrealistic to suppose that when Plato embarked on his career as a writer, he made a conscious decision to put all of the compositions that he would henceforth compose for a general reading public (with the exception of *Apology*) in the form of a dialogue. If the question, "why did Plato write dialogues?", which many of his readers are tempted to ask, pre-supposes that there must have been some such once-and-for-all decision, then it is poorly posed. It makes better sense to break that question apart into many little ones: better to ask, "Why did Plato write *this particular work* (for example: *Protagoras*, or *Republic*, or *Symposium*, or *Laws*) in the form of a dialogue—and *that* one *(Timaeus*, say) mostly in the form of a long and rhetorically elaborate single speech?" than to ask why he decided to adopt the dialogue form.

The best way to form a reasonable conjecture about why Plato wrote any given work in the form of a dialogue is to ask: what would be lost, were one to attempt to re-write this work in a way that eliminated the give-and-take of interchange, stripped the characters of their personality and social markers, and transformed the result into something that comes straight from the mouth of its author? This is often a question that will be easy to

answer, but the answer might vary greatly from one dialogue to another. In pursuing this strategy, we must not rule out the possibility that some of Plato's reasons for writing this or that work in the form of a dialogue will also be his reason for doing so in other cases—perhaps some of his reasons, so far as we can guess at them, will be present in all other cases. For example, the use of character and conversation allows an author to enliven his work, to awaken the interest of his readership, and therefore to reach a wider audience. The enormous appeal of Plato's writings is in part a result of their dramatic composition. Even treatise-like compositions —*Timaeus* and *Laws*, for example—improve in readability because of their conversational frame. Furthermore, the dialogue form allows Plato's evident interest in pedagogical questions (how is it possible to learn? what is the best way to learn? from what sort of person can we learn? what sort of person is in a position to learn?) to be pursued not only in the content of his compositions but also in their form. Even in *Laws* such questions are not far from Plato's mind, as he demonstrates, through the dialogue form, how it is possible for the citizens of Athens, Sparta, and Crete to learn from each other by adapting and improving upon each other's social and political institutions.

In some of his works, it is evident that one of Plato's goals is to create a sense of puzzlement among his readers, and that the dialogue form is being used for this purpose. The *Parmenides* is perhaps the clearest example of such a work, because here Plato relentlessly rubs his readers' faces in a baffling series of unresolved puzzles and apparent contradictions. But several of his other works also have this character, though to a smaller degree: for example, *Protagoras* (can virtue be taught?), *Hippias Minor* (is voluntary wrongdoing better than involuntary wrongdoing?), and portions of *Meno* (are some people virtuous because of divine inspiration?). Just as someone who encounters Socrates in conversation should sometimes be puzzled about whether he means what he says (or whether he is instead speaking ironically), so Plato sometimes

uses the dialogue form to create in his readers a similar sense of discomfort about what he means and what we ought to infer from the arguments that have been presented to us. But Socrates does not *always* speak ironically, and similarly Plato's dialogues do not *always* aim at creating a sense of bafflement about what we are to think about the subject under discussion. There is no mechanical rule for discovering how best to read a dialogue, no interpretive strategy that applies equally well to all of his works. We will best understand Plato's works and profit most from our reading of them if we recognize their great diversity of styles and adapt our way of reading accordingly. Rather than impose on our reading of Plato a uniform expectation of what he must be doing (because he has done such a thing elsewhere), we should bring to each dialogue a receptivity to what is unique to it. That would be the most fitting reaction to the artistry in his philosophy.

Bibliography

The bibliography below is meant as a highly selective and limited guide for readers who want to learn more about the issues covered above. Further discussion of these and other issues regarding Plato's philosophy, and far more bibliographical information, is available in the other entries on Plato.

Translations into English

Cooper, John M. (ed.), 1997, *Plato: Complete Works*, Indianapolis: Hackett. (Contains translations of all the works handed down from antiquity with attribution to Plato, some of which are universally agreed to be spurious, with explanatory footnotes and both a general Introduction to the study of the dialogues and individual Introductory Notes to each work translated.)

Aristotle

Aristotle (384–322 B.C.E.) numbers among the greatest philosophers of all time. Judged solely in terms of his philosophical influence, only Plato is his peer: Aristotle's works shaped centuries of philosophy from Late Antiquity through the Renaissance, and even today continue to be studied with keen, non-antiquarian interest. A prodigious researcher and writer, Aristotle left a great body of work, perhaps numbering as many as two-hundred treatises, from which approximately thirty-one survive.[1] His extant writings span a wide range of disciplines, from logic, metaphysics and philosophy of mind, through ethics, political theory, aesthetics and rhetoric, and into such primarily non-philosophical fields as empirical biology, where he excelled at detailed plant and animal observation and description. In all these areas, Aristotle's theories have provided illumination, met with resistance, sparked debate, and generally stimulated the sustained interest of an abiding readership.

Because of its wide range and its remoteness in time, Aristotle's philosophy defies easy encapsulation. The long history of interpretation and appropriation of Aristotelian texts and themes—spanning over two millennia and comprising philosophers working within a variety of religious and secular traditions—has rendered even basic points of interpretation controversial. The set of entries on Aristotle in this site addresses this situation by proceeding in three tiers. First, the present, general entry offers a brief account of Aristotle's life and characterizes his central philosophical commitments, highlighting his most distinctive methods and most influential achievements.[2] Second are *General Topics*, which offer detailed introductions to the main areas of Aristotle's philosophical activity. Finally, there follow *Special Topics*, which

investigate in greater detail more narrowly focused issues, especially those of central concern in recent Aristotelian scholarship.

1. Aristotle's Life

Born in 384 B.C.E. in the Macedonian region of northeastern Greece in the small city of Stagira (whence the moniker 'the Stagirite'), Aristotle was sent to Athens at about the age of seventeen to study in Plato's Academy, then a pre-eminent place of learning in the Greek world. Once in Athens, Aristotle remained associated with the Academy until Plato's death in 347, at which time he left for Assos, in Asia Minor, on the northwest coast of present-day Turkey. There he continued the philosophical activity he had begun in the Academy, but in all likelihood also began to expand his researches into marine biology. He remained at Assos for approximately three years, when, evidently upon the death of his host Hermeias, a friend and former Academic who had been the ruler of Assos, Aristotle moved to the nearby coastal island of Lesbos. There he continued his philosophical and empirical researches for an additional two years, working in conjunction with Theophrastus, a native of Lesbos who was also reported in antiquity to have been associated with Plato's Academy. While in Lesbos, Aristotle married Pythias, the niece of Hermeias, with whom he had a daughter, also named Pythias.

In 343, upon the request of Philip, the king of Macedon, Aristotle left Lesbos for Pella, the Macedonian capital, in order to tutor the king's thirteen-year-old son, Alexander—the boy who was eventually to become Alexander the Great. Although speculation concerning Aristotle's influence upon the developing Alexander has proven irresistible to historians, in fact little concrete is known about their interaction. On the balance, it seems reasonable to conclude that some tuition took place, but that it lasted only two or three years, when Alexander was aged from thirteen to fifteen. By fifteen, Alexander was apparently already serving as a deputy military commander for his father, a circumstance undermining, if inconclusively, the judgment of those historians who conjecture a longer

period of tuition. Be that as it may, some suppose that their association lasted as long as eight years.

It is difficult to rule out that possibility decisively, since little is known about the period of Aristotle's life from 341–335. He evidently remained a further five years in Stagira or Macedon before returning to Athens for the second and final time, in 335. In Athens, Aristotle set up his own school in a public exercise area dedicated to the god Apollo Lykeios, whence its name, the *Lyceum*. Those affiliated with Aristotle's school later came to be called *Peripatetics*, probably because of the existence of an ambulatory (*peripatos*) on the school's property adjacent to the exercise ground. Members of the Lyceum conducted research into a wide range of subjects, all of which were of interest to Aristotle himself: botany, biology, logic, music, mathematics, astronomy, medicine, cosmology, physics, the history of philosophy, metaphysics, psychology, ethics, theology, rhetoric, political history, government and political theory, rhetoric, and the arts. In all these areas, the Lyceum collected manuscripts, thereby, according to some ancient accounts, assembling the first great library of antiquity.

During this period, Aristotle's wife, Pythias, died and he developed a new relationship with Herpyllis, perhaps like him a native of Stagira, though her origins are disputed, as is the question of her exact relationship to Aristotle. Some suppose that she was merely his slave; others infer from the provisions of Aristotle's will that she was a freed woman and likely his wife at the time of his death. In any event, they had children together, including a son, Nicomachus, named for Aristotle's father and after whom his *Nicomachean Ethics* is presumably named.

After thirteen years in Athens, Aristotle once again found cause to retire from the city, in 323. Probably his departure was occasioned by a resurgence of the always-simmering anti-Macedonian sentiment in Athens, which was free to come to the boil after Alexander succumbed to

disease in Babylon during that same year. Because of his connections to Macedon, Aristotle reasonably feared for his safety and left Athens, remarking, as an oft-repeated ancient tale would tell it, that he saw no reason to permit Athens to sin twice against philosophy. He withdrew directly to Chalcis, on Euboea, an island off the Attic coast, and died there of natural causes the following year, in 322.[3]

2. The Aristotelian Corpus: Character and Primary Divisions

Aristotle's writings tend to present formidable difficulties to his novice readers. To begin, he makes heavy use of unexplained technical terminology, and his sentence structure can at times prove frustrating. Further, on occasion a chapter or even a full treatise coming down to us under his name appears haphazardly organized, if organized at all; indeed, in several cases, scholars dispute whether a continuous treatise currently arranged under a single title was ever intended by Aristotle to be published in its present form or was rather stitched together by some later editor employing whatever principles of organization he deemed suitable.[4] This helps explain why students who turn to Aristotle after first being introduced to the supple and mellifluous prose on display in Plato's dialogues often find the experience frustrating. Aristotle's prose requires some acclimatization.

All the more puzzling, then, is Cicero's observation that if Plato's prose was silver, Aristotle's was a flowing river of gold (*Ac. Pr.* 38.119, cf. *Top.* 1.3, *De or.* 1.2.49). Cicero was arguably the greatest prose stylist of Latin and was also without question an accomplished and fair-minded critic of the prose styles of others writing in both Latin and Greek. We must assume, then, that Cicero had before him works of Aristotle other than those we possess. In fact, we know that Aristotle wrote dialogues, presumably while still in the Academy, and in their few surviving

remnants we are afforded a glimpse of the style Cicero describes. In most of what we possess, unfortunately, we find work of a much less polished character. Rather, Aristotle's extant works read like what they very probably are: lecture notes, drafts first written and then reworked, ongoing records of continuing investigations, and, generally speaking, in-house compilations intended not for a general audience but for an inner circle of auditors. These are to be contrasted with the "exoteric" writings Aristotle sometimes mentions, his more graceful compositions intended for a wider audience (*Pol*. 1278b30; *EE* 1217b22, 1218b34). Unfortunately, then, we are left for the most part, though certainly not entirely, with unfinished works in progress rather than with finished and polished productions. Still, many of those who persist with Aristotle come to appreciate the unembellished directness of his style.

More importantly, the unvarnished condition of Aristotle's surviving treatises does not hamper our ability to glean their philosophical content. His thirty-one surviving works (that is, those contained in the "Corpus Aristotelicum" of our medieval manuscripts that are judged to be authentic) all contain recognizably Aristotelian doctrine; and most of these contain theses whose basic purport is clear, even where matters of detail and nuance are subject to exegetical controversy.

These works may be categorized in terms of the intuitive organizational principles preferred by Aristotle. He refers to the branches of learning as "sciences" (*epistêmai*), best regarded as organized bodies of learning completed for presentation rather than as ongoing records of empirical researches. Moreover, again in his terminology, *natural sciences* such as physics are but one branch of *theoretical science*, which comprises both empirical and non-empirical pursuits. He distinguishes theoretical science from more practically oriented studies, some of which concern human conduct and others of which focus on the productive crafts. Thus, the Aristotelian sciences divide into three: (i) theoretical, (ii) practical, and

(iii) productive. The principles of division are straightforward: theoretical science seeks knowledge for its own sake; practical science concerns conduct and goodness in action, both individual and societal; and productive science aims at the creation of beautiful or useful objects (*Top*. 145a15–16; *Phys*. 192b8–12; *DC* 298a27–32, *DA* 403a27–b2; *Met*. 1025b25, 1026a18–19, 1064a16–19, b1–3; *EN* 1139a26–28, 1141b29–32).

(i) The *theoretical sciences* include prominently what Aristotle calls *first philosophy*, or metaphysics as we now call it, but also *mathematics*, and *physics*, or natural philosophy. Physics studies the natural universe as a whole, and tends in Aristotle's hands to concentrate on conceptual puzzles pertaining to nature rather than on empirical research; but it reaches further, so that it includes also a theory of causal explanation and finally even a proof of an unmoved mover thought to be the first and final cause of all motion. Many of the puzzles of primary concern to Aristotle have proven perennially attractive to philosophers, mathematicians, and theoretically inclined natural scientists. They include, as a small sample, Zeno's paradoxes of motion, puzzles about time, the nature of place, and difficulties encountered in thought about the infinite.

Natural philosophy also incorporates the special sciences, including biology, botany, and astronomical theory. Most contemporary critics think that Aristotle treats psychology as a sub-branch of natural philosophy, because he regards the soul (*psuchê*) as the basic principle of life, including all animal and plant life. In fact, however, the evidence for this conclusion is scanty. It is instructive to note that earlier periods of Aristotelian scholarship thought this controversial, so that, for instance, even something as innocuous-sounding as the question of the proper home of psychology in Aristotle's division of the sciences ignited a multi-decade debate in the Renaissance.[5]

(ii) *Practical sciences* are less contentious, at least as regards their range. These deal with conduct and action, both individual and societal. Practical science thus contrasts with theoretical science, which seeks knowledge for its own sake, and, less obviously, with the productive sciences, which deal with the creation of products external to sciences themselves. Both politics and ethics fall under this branch.

(iii) Finally, then, the *productive sciences* are mainly crafts aimed at the production of artefacts, or of human productions more broadly construed. The productive sciences include, among others, ship-building, agriculture, and medicine, but also the arts of music, theatre, and dance. Another form of productive science is rhetoric, which treats the principles of speech-making appropriate to various forensic and persuasive settings, including centrally political assemblies.

Significantly, Aristotle's tri-fold division of the sciences makes no mention of logic. Although he did not use the word 'logic' in our sense of the term, Aristotle in fact developed the first formalized system of logic and valid inference. In Aristotle's framework—although he is nowhere explicit about this—logic belongs to no one science, but rather formulates the principles of correct argumentation suitable to all areas of inquiry in common. It systematizes the principles licensing acceptable inference, and helps to highlight at an abstract level seductive patterns of incorrect inference to be avoided by anyone with a primary interest in truth. So, alongside his more technical work in logic and logical theory, Aristotle investigates informal styles of argumentation and seeks to expose common patterns of fallacious reasoning.

Aristotle's investigations into logic and the forms of argumentation make up part of the group of works coming down to us from the Middle Ages under the heading the *Organon* (*organon = tool* in Greek). Although not so characterized in these terms by Aristotle, the name is apt, so long as it

is borne in mind that intellectual inquiry requires a broad range of tools. Thus, in addition to logic and argumentation (treated primarily in the *Prior Analytics* and *Topics*), the works included in the *Organon* deal with category theory, the doctrine of propositions and terms, the structure of scientific theory, and to some extent the basic principles of epistemology.

When we slot Aristotle's most important surviving authentic works into this scheme, we end up with the following basic divisions of his major writings:

- Organon
    - *Categories* (*Cat.*)
    - *De Interpretatione* (*DI*) [*On Interpretation*]
    - *Prior Analytics* (*APr*)
    - *Posterior Analytics* (*APo*)
    - *Topics* (*Top.*)
    - *Sophistical Refutations* (*SE*)
- Theoretical Sciences
    - *Physics* (*Phys.*)
    - *Generation and Corruption* (*Gen. et Corr.*)
    - *De Caelo* (*DC*) [*On the Heavens*]
    - *Metaphysics* (*Met.*)
    - *De Anima* (*DA*) [*On the Soul*]
    - *Parva Naturalia* (*PN*) [*Brief Natural Treatises*]
    - *History of Animals* (*HA*)
    - *Parts of Animals* (*PA*)
    - *Movement of Animals* (*MA*)
    - *Meteorology* (*Meteor.*)
    - *Progression of Animals* (*IA*)
    - *Generation of Animals* (*GA*)
- Practical Sciences
    - *Nicomachean Ethics* (*EN*)

- *Eudemian Ethics* (*EE*)
- *Magna Moralia* (*MM*) [*Great Ethics*]
- *Politics* (*Pol.*)
- Productive Science
  - *Rhetoric* (*Rhet.*)
  - *Poetics* (*Poet.*)

The titles in this list are those in most common use today in English-language scholarship, followed by standard abbreviations in parentheses. For no discernible reason, Latin titles are customarily employed in some cases, English in others. Where Latin titles are in general use, English equivalents are given in square brackets.

3. *Phainomena* and the Endoxic Method

Aristotle's basic approach to philosophy is best grasped initially by way of contrast. Whereas Descartes seeks to place philosophy and science on firm foundations by subjecting all knowledge claims to a searing methodological doubt, Aristotle begins with the conviction that our perceptual and cognitive faculties are basically dependable, that they for the most part put us into direct contact with the features and divisions of our world, and that we need not dally with sceptical postures before engaging in substantive philosophy. Accordingly, he proceeds in all areas of inquiry in the manner of a modern-day natural scientist, who takes it for granted that progress follows the assiduous application of a well-trained mind and so, when presented with a problem, simply goes to work. When he goes to work, Aristotle begins by considering how the world appears, reflecting on the puzzles those appearances throw up, and reviewing what has been said about those puzzles to date. These methods comprise his twin appeals to *phainomena* and the endoxic method.

These two methods reflect in different ways Aristotle's deepest motivations for doing philosophy in the first place. "Human beings began to do philosophy," he says, "even as they do now, because of wonder, at first because they wondered about the strange things right in front of them, and then later, advancing little by little, because they came to find greater things puzzling" (*Met*. 982b12). Human beings philosophize, according to Aristotle, because they find aspects of their experience puzzling. The sorts of puzzles we encounter in thinking about the universe and our place within it—*aporiai*, in Aristotle's terminology—tax our understanding and induce us to philosophize.

According to Aristotle, it behooves us to begin philosophizing by laying out the *phainomena*, the *appearances*, or, more fully, *the things appearing to be the case*, and then also collecting the *endoxa*, the credible opinions handed down regarding matters we find puzzling. As a typical example, in a passage of his *Nicomachean Ethics*, Aristotle confronts a puzzle of human conduct, the fact that we are apparently sometimes akratic or weak-willed. When introducing this puzzle, Aristotle pauses to reflect upon a precept governing his approach to philosophy:

> As in other cases, we must set out the appearances (*phainomena*) and run through all the puzzles regarding them. In this way we must prove the credible opinions (*endoxa*) about these sorts of experiences—ideally, all the credible opinions, but if not all, then most of them, those which are the most important. For if the objections are answered and the credible opinions remain, we shall have an adequate proof. (*EN* 1145b2–7)

Scholars dispute concerning the degree to which Aristotle regards himself as beholden to the credible opinions (*endoxa*) he recounts and the basic appearances (*phainomena*) to which he appeals.[6] Of course, since the *endoxa* will sometimes conflict with one another, often precisely because

the *phainomena* generate *aporiai*, or puzzles, it is not always possible to respect them in their entirety. So, as a group they must be re-interpreted and systematized, and, where that does not suffice, some must be rejected outright. It is in any case abundantly clear that Aristotle is willing to abandon some or all of the *endoxa* and *phainomena* whenever science or philosophy demands that he do so (*Met*. 1073b36, 1074b6; *PA* 644b5; *EN* 1145b2–30).

Still, his attitude towards *phainomena* does betray a preference to conserve as many appearances as is practicable in a given domain—not because the appearances are unassailably accurate, but rather because, as he supposes, appearances tend to track the truth. We are outfitted with sense organs and powers of mind so structured as to put us into contact with the world and thus to provide us with data regarding its basic constituents and divisions. While our faculties are not infallible, neither are they systematically deceptive or misdirecting. Since philosophy's aim is truth and much of what appears to us proves upon analysis to be correct, *phainomena* provide both an impetus to philosophize and a check on some of its more extravagant impulses.

Of course, it is not always clear what constitutes a *phainomenon*; still less is it clear which *phainomenon* is to be respected in the face of *bona fide* disagreement. This is in part why Aristotle endorses his second and related methodological precept, that we ought to begin philosophical discussions by collecting the most stable and entrenched opinions regarding the topic of inquiry handed down to us by our predecessors. Aristotle's term for these privileged views, *endoxa*, is variously rendered as 'reputable opinions', 'credible opinions', 'entrenched beliefs', 'credible beliefs', or 'common beliefs'. Each of these translations captures at least part of what Aristotle intends with this word, but it is important to appreciate that it is a fairly technical term for him. An *endoxon* is the sort of opinion we spontaneously regard as reputable or worthy of respect, even if upon

reflection we may come to question its veracity. (Aristotle appropriates this term from ordinary Greek, in which an *endoxos* is a notable or honourable man, a man of high repute whom we would spontaneously respect—though we might, of course, upon closer inspection, find cause to criticize him.) As he explains his use of the term, *endoxa* are widely shared opinions, often ultimately issuing from those we esteem most: '*Endoxa* are those opinions accepted by everyone, or by the majority, or by the wise—and among the wise, by all or most of them, or by those who are the most notable and having the highest reputation' (*Top*. 100b21–23). *Endoxa* play a special role in Aristotelian philosophy in part because they form a significant sub-class of *phainomena* (*EN* 1154b3–8): because they are the privileged opinions we find ourselves unreflectively endorsing and reaffirming after some reflection, they themselves come to qualify as appearances to be preserved where possible.

For this reason, Aristotle's method of beginning with the *endoxa* is more than a pious platitude to the effect that it behooves us to mind our superiors. He does think this, as far as it goes, but he also maintains, more instructively, that we can be led astray by the terms within which philosophical problems are bequeathed to us. Very often, the puzzles confronting us were given crisp formulations by earlier thinkers and we find them puzzling precisely for that reason. Equally often, however, if we reflect upon the terms within which the puzzles are cast, we find a way forward; when a formulation of a puzzle betrays an untenable structuring assumption, a solution naturally commends itself. This is why in more abstract domains of inquiry we are likely to find ourselves seeking guidance from our predecessors even as we call into question their ways of articulating the problems we are confronting.

Aristotle applies his method of running through the *phainomena* and collecting the *endoxa* widely, in nearly every area of his philosophy. To take a typical illustration, we find the method clearly deployed in his

discussion of time in *Physics* iv 10–14. We begin with a *phainomenon*: we feel sure *that time exists* or at least *that time passes*. So much is, inescapably, how our world appears: we experience time as passing, as unidirectional, as unrecoverable when lost. Yet when we move to offer an account of what time might be, we find ourselves flummoxed. For guidance, we turn to what has been said about time by those who have reflected upon its nature. It emerges directly that both philosophers and natural scientists have raised problems about time.

As Aristotle sets them out, these problems take the form of puzzles, or *aporiai*, regarding whether and if so how time exists (*Phys*. 218a8–30). If we say that time is the totality of the past, present and future, we immediately find someone objecting that time exists but that the past and future do not. According to the objector, only the present exists. If we retort then that time is what *did* exist, what exists *at present* and what *will* exist, then we notice first that our account is insufficient: after all, there are many things which did, do, or will exist, but these are things that are *in* time and so not the same as time itself. We further see that our account already threatens circularity, since to say that something *did* or *will* exist seems only to say that it existed at an earlier *time* or will come to exist at a later *time*. Then again we find someone objecting to our account that even the notion of the *present* is troubling. After all, either the present is constantly changing or it remains forever the same. If it remains forever the same, then the current present is the same as the present of 10,000 years ago; yet that is absurd. If it is constantly changing, then no two presents are the same, in which case a past present must have come into and out of existence before the present present. When? Either it went out of existence even as it came into existence, which seems odd to say the least, or it went out of existence at some instant after it came into existence, in which case, again, two presents must have existed at the same instant.

In setting such *aporiai*, Aristotle does not mean to endorse any given *endoxon* on one side or the other. Rather, he thinks that such considerations present credible puzzles, reflection upon which may steer us towards a deeper understanding of the nature of time. In this way, *aporiai* bring into sharp relief the issues requiring attention if progress is to be made. Thus, by reflecting upon the *aporiai* regarding time, we are led immediately to think about duration and divisibility, about *quanta* and *continua*, and about a variety of categorial questions. That is, if time exists, then what sort of thing is it? Is it the sort of thing which exists absolutely and independently? Or is it rather the sort of thing which, like a surface, depends upon other things for its existence? When we begin to address these sorts of questions, we also begin to ascertain the sorts of assumptions at play in the *endoxa* coming down to us regarding the nature of time. Consequently, when we collect the *endoxa* and survey them critically, we learn something about our quarry, in this case about the nature of time—and crucially also something about the constellation of concepts which must be refined if we are to make genuine philosophical progress with respect to it. What holds in the case of time, contends Aristotle, holds generally. This is why he characteristically begins a philosophical inquiry by presenting the *phainomena*, collecting the *endoxa*, and running through the puzzles to which they give rise.

4. Logic, Science, and Dialectic

Aristotle's reliance on *endoxa* takes on a still greater significance given the role such opinions play in *dialectic*, which he regards as an important form of non-scientific reasoning. Dialectic, like science (*epistêmê*), trades in logical inference; but science requires premises of a sort beyond the scope of ordinary dialectical reasoning. Whereas science relies upon premises which are necessary and known to be so, a dialectical discussion can proceed by relying on *endoxa*, and so can claim only to be as secure as the

endoxa upon which it relies. This is not a problem, suggests Aristotle, since we often reason fruitfully and well in circumstances where we cannot claim to have attained scientific understanding. Minimally, however, all reasoning—whether scientific or dialectical—must respect the canons of logic and inference.

4.1 Logic

Among the great achievements to which Aristotle can lay claim is the first systematic treatment of the principles of correct reasoning, the first logic. Although today we recognize many forms of logic beyond Aristotle's, it remains true that he not only developed a theory of deduction, now called syllogistic, but added to it a modal syllogistic and went a long way towards proving some meta-theorems pertinent to these systems. Of course, philosophers before Aristotle reasoned well or reasoned poorly, and the competent among them had a secure working grasp of the principles of validity and soundness in argumentation. No-one before Aristotle, however, developed a systematic treatment of the principles governing correct inference; and no-one before him attempted to codify the formal and syntactic principles at play in such inference. Aristotle somewhat uncharacteristically draws attention to this fact at the end of a discussion of logic inference and fallacy:

> Once you have surveyed our work, if it seems to you that our system has developed adequately in comparison with other treatments arising from the tradition to date—bearing in mind how things were at the beginning of our inquiry—it falls to you, our students, to be indulgent with respect to any omissions in our system, and to feel a great debt of gratitude for the discoveries it contains. (*Soph. Ref.* 184b2–8)

Even if we now regard it as commonplace that his logic is but a fraction of the logic we know and use, Aristotle's accomplishment was so encompassing that no less a figure than Kant, writing over two millennia after the appearance of Aristotle's treatises on logic, found it easy to offer an appropriately laudatory judgment: 'That from the earliest times logic has traveled a secure course can be seen from the fact that since the time of Aristotle it has not had to go a single step backwards…What is further remarkable about logic is that until now it has also been unable to take a single step forward, and therefore seems to all appearance to be finished and complete' (*Critique of Pure Reason* B vii).

In Aristotle's logic, the basic ingredients of reasoning are given in terms of *inclusion* and *exclusion* relations, of the sort graphically captured many years later by the device of Venn diagrams. He begins with the notion of a patently correct sort of argument, one whose evident and unassailable acceptability induces Aristotle to refer to is as a 'perfect deduction' (*APr*. 24b22–25). Generally, a *deduction* (*sullogismon*), according to Aristotle, is a valid or acceptable argument. More exactly, a deduction is 'an argument in which when certain things are laid down something else follows of necessity in virtue of their being so' (*APr*. 24b18–20). His view of deductions is, then, akin to a notion of validity, though there are some minor differences. For example, Aristotle maintains that irrelevant premises will ruin a deduction, whereas validity is indifferent to irrelevance or indeed to the addition of premises of any kind to an already valid argument. Moreover, Aristotle insists that deductions make progress, whereas every inference from p to p is trivially valid. Still, Aristotle's general conception of deduction is sufficiently close to validity that we may pass into speaking in terms of valid structures when characterizing his syllogistic. In general, he contends that a deduction is the sort of argument whose *structure* guarantees its validity, irrespective of the truth or falsity of its premises. This holds intuitively for the following structure:

1. All *A*s are *B*s.
2. All *B*s are *C*s.
3. Hence, all *A*s are *C*s.

Accordingly, anything taking this form will be a deduction in Aristotle's sense. Let the *A*s, *B*s, and *C*s be anything at all, and *if* indeed the *A*s are *B*s, and the *B*s *C*s, then *of necessity* the *A*s will be *C*s. This particular deduction is *perfect* because its validity needs no proof, and perhaps because it admits of no proof either: any proof would seem to rely ultimately upon the intuitive validity of this sort of argument.

Aristotle seeks to exploit the intuitive validity of perfect deductions in a surprisingly bold way, given the infancy of his subject: he thinks he can establish principles of transformation in terms of which *every* deduction (or, more precisely, every non-modal deduction) can be translated into a perfect deduction. He contends that by using such transformations we can place all deduction on a firm footing.

If we focus on just the simplest kinds of deduction, Aristotle's procedure comes quickly into view. The perfect deduction already presented is an instance of universal affirmation: all *A*s are *B*s; all *B*s *C*s; and so, all *A*s are *C*s. Now, contends Aristotle, it is possible to run through all combinations of simple premises and display their basic inferential structures and then to relate them back to this and similarly perfect deductions. Thus, if we vary the *quantity* of a proposition's subject (universal *all* versus indeterminate *some*) along with the *quality* or kind of the predication (*positive* versus *negative*), we arrive at all the possible combinations of the most basic kind of arguments.

It turns out that some of these arguments are deductions, or valid syllogisms, and some are not. Those which are not admit of counterexamples, whereas those which are, of course, do not. There are

counterexamples to those, for instance, suffering from what came to be called undistributed middle terms, e.g.: all *A*s are *B*s; some *B*s are *C*s; so, all *A*s are *C*s (all university students are literate; some literate people read poetry; so, all university students read poetry). There is no counterexample to the perfect deduction in the form of a universal affirmation: if all *A*s are *B*s, and all *B*s *C*s, then there is no escaping the fact that all *A*s are *C*s. So, if all the kinds of deductions possible can be reduced to the intuitively valid sorts, then the validity of all can be vouchsafed.

To effect this sort of reduction, Aristotle relies upon a series of meta-theorems, some of which he proves and others of which he merely reports (though it turns out that they do all indeed admit of proofs). His principles are *meta*-theorems in the sense that no argument can run afoul of them and still qualify as a genuine deduction. They include such theorems as: (i) no deduction contains two negative premises; (ii) a deduction with a negative conclusion must have a negative premise; (iii) a deduction with a universal conclusion requires two universal premises; and (iv) a deduction with a negative conclusion requires exactly one negative premise. He does, in fact, offer proofs for the most significant of his meta-theorems, so that we can be assured that all deductions in his system are valid, even when their validity is difficult to grasp immediately.

In developing and proving these meta-theorems of logic, Aristotle charts territory left unexplored before him and unimproved for many centuries after his death.

For a fuller account of Aristotle's achievements in logic, see the entry on Aristotle's Logic.

4.2 Science

Aristotle approaches the study of logic not as an end in itself, but with a view to its role in human inquiry and explanation. Logic is a tool, he

thinks, one making an important but incomplete contribution to science and dialectic. Its contribution is incomplete because science (*epistêmê*) employs arguments which are more than mere deductions. A deduction is minimally a valid syllogism, and certainly science must employ arguments passing this threshold. Still, science needs more: a science proceeds by *organizing* the data in its domain into a series of arguments which, beyond being deductions, feature premises which are necessary and, as Aristotle says, "better known by nature", or "more intelligible by nature" (*gnôrimôteron phusei*) (*APo*. 71b33–72a25; *Top*. 141b3–14; *Phys*. 184a16–23). By this he means that they should reveal the genuine, mind-independent natures of things.

He further insists that science (*epistêmê*)—a comparatively broad term in his usage, since it extends to fields of inquiry like mathematics and metaphysics no less than the empirical sciences—not only reports the facts but also explains them by displaying their priority relations (*APo*. 78a22–28). That is, science explains what is less well known by what is better known and more fundamental, and what is explanatorily anemic by what is explanatorily fruitful.

We may, for instance, wish to know why trees lose their leaves in the autumn. We may say, rightly, that this is due to the wind blowing through them. Still, this is not a deep or general explanation, since the wind blows equally at other times of year without the same result. A deeper explanation—one unavailable to Aristotle but illustrating his view nicely—is more general, and also more causal in character: trees shed their leaves because diminished sunlight in the autumn inhibits the production of chlorophyll, which is required for photosynthesis, and without photosynthesis trees go dormant. Importantly, science should not only record these facts but also display them in their correct explanatory order. That is, although a deciduous tree which fails to photosynthesize is also a tree lacking in chlorophyll production, its failing to produce chlorophyll

explains its inability to photosynthesize and not the other way around. This sort of asymmetry must be captured in scientific explanation. Aristotle's method of scientific exposition is designed precisely to discharge this requirement.

Science seeks to capture not only the causal asymmetries in nature, but also its deep, invariant patterns. Consequently, in addition to being explanatorily basic, the first premise in a scientific deduction will be necessary. So, says Aristotle:

> We think we understand a thing without qualification, and not in the sophistic, accidental way, whenever we think we know the cause in virtue of which something is—that it is the cause of that very thing—*and also* know that this cannot be otherwise. Clearly, knowledge (*epistêmê*) is something of this sort. After all, both those with knowledge and those without it suppose that this is so—although only those with knowledge are actually in this condition. Hence, whatever is known without qualification cannot be otherwise. (*APo* 71b9–16; cf. *APo* 71b33–72a5; *Top.* 141b3–14, *Phys.* 184a10–23; *Met.* 1029b3–13)

For this reason, science requires more than mere deduction. Altogether, then, the currency of science is *demonstration* (*apodeixis*), where a demonstration is a deduction with premises revealing the causal structures of the world, set forth so as to capture what is necessary and to reveal what is better known and more intelligible by nature (*APo* 71b33–72a5, *Phys.* 184a16–23, *EN* 1095b2–4).

Aristotle's approach to the appropriate form of scientific explanation invites reflection upon a troubling epistemological question: how does demonstration begin? If we are to lay out demonstrations such that the less well known is inferred by means of deduction from the better known, then

unless we reach rock-bottom, we will evidently be forced either to continue ever backwards towards the increasingly better known, which seems implausibly endless, or lapse into some form of circularity, which seems undesirable. The alternative seems to be permanent ignorance. Aristotle contends:

> Some people think that since knowledge obtained via demonstration requires the knowledge of primary things, there is no knowledge. Others think that there is knowledge and that all knowledge is demonstrable. Neither of these views is either true or necessary. The first group, those supposing that there is no knowledge at all, contend that we are confronted with an infinite regress. They contend that we cannot know posterior things because of prior things if none of the prior things is primary. Here what they contend is correct: it is indeed impossible to traverse an infinite series. Yet, they maintain, if the regress comes to a halt, and there are first principles, they will be unknowable, since surely there will be no demonstration of first principles—given, as they maintain, that only what is demonstrated can be known. But if it is not possible to know the primary things, then neither can we know without qualification or in any proper way the things derived from them. Rather, we can know them instead only on the basis of a hypothesis, to wit, *if* the primary things obtain, then so too do the things derived from them. The other group agrees that knowledge results only from demonstration, but believes that nothing stands in the way of demonstration, since they admit circular and reciprocal demonstration as possible. (*APo*. 72b5–21)

Aristotle's own preferred alternative is clear:

> We contend that not all knowledge is demonstrative: knowledge of the immediate premises is indemonstrable. Indeed, the necessity

here is apparent; for if it is necessary to know the prior things, that is, those things from which the demonstration is derived, and if eventually the regress comes to a standstill, it is necessary that these immediate premises be indemonstrable. (*APo*. 72b21–23)

In sum, if all knowledge requires demonstration, and all demonstration proceeds from what is more intelligible by nature to what is less so, then either the process goes on indefinitely or it comes to a halt in undemonstrated first principles, which are known, and known securely. Aristotle dismisses the only remaining possibility, that demonstration might be circular, rather curtly, with the remark that this amounts to 'simply saying that something is the case if it is the case,' by which device 'it is easy to prove anything' (*APo*. 72b32–73a6).

Aristotle's own preferred alternative, that there are first principles of the sciences graspable by those willing to engage in assiduous study, has caused consternation in many of his readers. In *Posterior Analytics* ii 19, he describes the process by which knowers move from perception to memory, and from memory to experience (*empeiria*)—which is a fairly technical term in this connection, reflecting the point at which a single universal comes to take root in the mind—and finally from experience to a grasp of first principles. This final intellectual state Aristotle characterizes as a kind of unmediated intellectual apprehension (*nous*) of first principles (*APo*. 100a10–b6).

Scholars have understandably queried what seems a casually asserted passage from the contingent, given in sense experience, to the necessary, as required for the first principles of science. Perhaps, however, Aristotle simply envisages a kind of *a posteriori* necessity for the sciences, including the natural sciences. In any event, he thinks that we can and do have knowledge, so that somehow we begin in sense perception and build up to an understanding of the necessary and invariant features of the

world. This is the knowledge featured in genuine science (*epistêmê*). In reflecting on the sort of progression Aristotle envisages, some commentators have charged him with an epistemological optimism bordering on the naïve; others contend that it is rather the charge of naïveté which is itself naïve, betraying as it does an unargued and untenable alignment of the necessary and the *a priori*.[7]

4.3 Dialectic

Not all rigorous reasoning qualifies as scientific. Indeed, little of Aristotle's extant writing conforms to the demands for scientific presentation laid down in the *Posterior Analytics*. As he recognizes, we often find ourselves reasoning from premises which have the status of *endoxa*, opinions widely believed or endorsed by the wise, even though they are not known to be necessary. Still less often do we reason having first secured the first principles of our domain of inquiry. So, we need some 'method by which we will be able to reason deductively about any matter proposed to us on the basis of *endoxa*, and to give an account of ourselves [when we are under examination by an interlocutor] without lapsing into contradiction' (*Top*. 100a18–20). This method he characterizes as *dialectic*.

The suggestion that we often use dialectic when engaged in philosophical exchange reflects Aristotle's supposition that there are two sorts of dialectic: one negative, or destructive, and the other positive, or constructive. In fact, in his work dedicated to dialectic, the *Topics*, he identifies three roles for dialectic in intellectual inquiry, the first of which is mainly preparatory:

> Dialectic is useful for three purposes: for training, for conversational exchange, and for sciences of a philosophical sort. That it is useful for training purposes is directly evident on the

basis of these considerations: once we have a direction for our inquiry we will more readily be able to engage a subject proposed to us. It is useful for conversational exchange because once we have enumerated the beliefs of the many, we shall engage them not on the basis of the convictions of others but on the basis of their own; and we shall re-orient them whenever they appear to have said something incorrect to us. It is useful for philosophical sorts of sciences because when we are able to run through the puzzles on both sides of an issue we more readily perceive what is true and what is false. Further, it is useful for uncovering what is primary among the commitments of a science. For it is impossible to say anything regarding the first principles of a science on the basis of the first principles proper to the very science under discussion, since among all the commitments of a science, the first principles are the primary ones. This comes rather, necessarily, from discussion of the credible beliefs (*endoxa*) belonging to the science. This is peculiar to dialectic, or is at least most proper to it. For since it is what cross-examines, dialectic contains the way to the first principles of all inquiries. (*Top*. 101a26–b4)

The first two of the three forms of dialectic identified by Aristotle are rather limited in scope. By contrast, the third is philosophically significant.

In its third guise, dialectic has a role to play in 'science conducted in a philosophical manner' (*pros tas kata philosphian epistêmas*; *Top*. 101a27–28, 101a34), where this sort of science includes what we actually find him pursuing in his major philosophical treatises. In these contexts, dialectic helps to sort the *endoxa*, relegating some to a disputed status while elevating others; it submits *endoxa* to cross-examination in order to test their staying power; and, most notably, according to Aristotle, dialectic puts us on the road to first principles (*Top*. 100a18–b4). If that is so, then dialectic plays a significant role in the order of philosophical discovery:

we come to establish first principles in part by determining which among our initial *endoxa* withstand sustained scrutiny. Here, as elsewhere in his philosophy, Aristotle evinces a noteworthy confidence in the powers of human reason and investigation.

5. Essentialism and Homonymy

However we arrive at secure principles in philosophy and science, whether by some process leading to a rational grasping of necessary truths, or by sustained dialectical investigation operating over judiciously selected *endoxa*, it does turn out, according to Aristotle, that we can uncover and come to know genuinely necessary features of reality. Such features, suggests Aristotle, are those captured in the essence-specifying definitions used in science (again in the broad sense of *epistêmê*).

Aristotle's commitment to essentialism runs deep. He relies upon a host of loosely related locutions when discussing the essences of things, and these give some clue to his general orientation. Among the locutions one finds rendered as *essence* in contemporary translations of Aristotle into English are: (i) *to ti esti* (the what it is); (ii) *to einai* (being); (iii) *ousia* (being); (iv) *hoper esti* (precisely what something is) and, most importantly, (v) *to ti ên einai* (the what it was to be) (*APo* 83a7; *Top*. 141b35; *Phys*. 190a17, 201a18–21; *Gen*. *et Corr*. 319b4; *DA* 424a25, 429b10; *Met*. 1003b24, 1006a32, 1006b13; *EN* 1102a30, 1130a12–13). Among these, the last locution (v) requires explication both because it is the most peculiar and because it is Aristotle's favored technical term for essence. It is an abbreviated way of saying 'that which it was for an instance of kind *K* to be an instance of kind *K*,' for instance 'that which it was (all along) for a human being to be a human being'. In speaking this way, Aristotle supposes that if we wish to know what a human being is, we cannot identify transient or non-universal features of that kind; nor indeed can we identify even universal features which do not run explanatorily deep.

Rather, as his preferred locution indicates, he is interested in what makes a human being human—and he assumes, first, that there is some feature F which all and only humans have in common and, second, that F explains the other features which we find across the range of humans.

Importantly, this second feature of Aristotelian essentialism differentiates his approach from the now more common modal approach, according to which:[8]

> F is an essential property of $x =_{df}$ if x loses F, then x ceases to exist.

Aristotle rejects this approach for several reasons, including most notably that he thinks that certain non-essential features satisfy the definition. Thus, beyond the categorical and logical features (everyone is such as to be either identical or not identical with the number nine), Aristotle recognizes a category of properties which he calls *idia* (*Cat*. 3a21, 4a10; *Top*. 102a18–30, 134a5–135b6), now usually known by their Medieval Latin rendering *propria*. *Propria* are non-essential properties which flow from the essence of a kind, such that they are necessary to that kind even without being essential. For instance, if we suppose that *being rational* is essential to human beings, then it will follow that every human being is *capable of grammar*. Being capable of grammar is not the same property as being rational, though it follows from it. Aristotle assumes his readers will appreciate that *being rational* asymmetrically explains *being capable of grammar*, even though, necessarily, something is rational if and only if it is also capable of grammar. Thus, because it is explanatorily prior, *being rational* has a better claim to being the essence of human beings than does *being capable of grammar*. Consequently, Aristotle's essentialism is more fine-grained than mere modal essentialism. Aristotelian essentialism holds:

> F is an essential property of $x =_{df}$ (i) if x loses F, then x ceases to

exist; and (ii) *F* is in an objective sense an explanatorily basic feature of *x*.

In sum, in Aristotle's approach, what it is to be, for instance, a human being is just what it always has been and always will be, namely *being rational*. Accordingly, this is the feature to be captured in an essence-specifying account of human beings (*APo* 75a42–b2; *Met*. 103b1–2, 1041a25–32).

Aristotle believes for a broad range of cases that kinds have essences discoverable by diligent research. He in fact does not devote much energy to arguing for this contention; still less is he inclined to expend energy combating anti-realist challenges to essentialism, perhaps in part because he is impressed by the deep regularities he finds, or thinks he finds, underwriting his results in biological investigation.[9] Still, he cannot be accused of profligacy regarding the prospects of essentialism.

On the contrary, he denies essentialism in many cases where others are prepared to embrace it. One finds this sort of denial prominently, though not exclusively, in his criticism of Plato. Indeed, it becomes a signature criticism of Plato and Platonists for Aristotle that many of their preferred examples of sameness and invariance in the world are actually cases of *multivocity*, or *homonymy* in his technical terminology. In the opening of the *Categories*, Aristotle distinguishes between *synonymy* and *homonymy* (later called *univocity* and *multivocity*). His preferred phrase for multivocity, which is extremely common in his writings, is 'being spoken of in many ways', or, more simply, 'multiply meant': *pollochôs legomenon*). All these locutions have a quasi-technical status for him. The least complex is univocity:

a and *b* are univocally *F* iff (i) *a* is *F*, (ii) *b* is *F*, and (iii) the accounts of *F*-ness in '*a* is *F*' and '*b* is *F*' are the same.

Thus, for instance, since the accounts of 'human' in 'Socrates is human' and 'Plato is human' will be the same, 'human' is univocal or synonymous in these applications. (Note that Aristotle's notion of synonymy is not the same as the contemporary English usage where it applies to *different* words with the same meaning.) In cases of univocity, we expect single, non-disjunctive definitions which capture and state the essence of the kinds in question. Let us allow once more for purposes of illustration that the essence-specifying definition of *human* is *rational animal*. Then, since *human* means *rational animal* across the range of its applications, there is some single essence to all members of the kind.

By contrast, when synonymy fails we have homonymy. According to Aristotle:

> *a* and *b* are homonymously *F* iff (i) *a* is *F*, (ii) *b* is *F*, (iii) the accounts of *F*-ness in '*a* is *F*' and '*b* is *F*' do not completely overlap.

To take an easy example without philosophical significance, *bank* is homonymous in 'Socrates and Alcibiades had a picnic on the bank' and 'Socrates and Alcibiades opened a joint account at the bank.' This case is illustrative, if uninteresting, because the accounts of *bank* in these occurrences have nothing whatsoever in common. Part of the interest in Aristotle's account of homonymy resides in its allowing partial overlap. Matters become more interesting if we examine whether—to use an illustration well suited to Aristotle's purposes but left largely unexplored by him—*conscious* is synonymous across 'Charlene was conscious of some awkwardness created by her remarks' and 'Higher vertebrates, unlike mollusks, are conscious.' In these instances, the situation with respect to synonymy or homonymy is perhaps not immediately clear, and so requires reflection and philosophical investigation.

Very regularly, according to Aristotle, this sort of reflection leads to an interesting discovery, namely that we have been presuming a univocal account where in fact none is forthcoming. This, according to Aristotle, is where the Platonists go wrong: they presume univocity where the world delivers homonymy or multivocity. (For a vivid illustration of Plato's univocity assumption at work, see *Meno* 71e1–72a5, where Socrates insists that there is but one kind of *excellence* (*aretê*) common to all kinds of excellent people, not a separate sort for men, women, slaves, children, and so on.) In one especially important example, Aristotle parts company with Plato over the univocity of goodness:

> We had perhaps better consider the universal good and run through the puzzles concerning what is meant by it—even though this sort of investigation is unwelcome to us, because those who introduced the Forms are friends of ours. Yet presumably it would be the better course to destroy even what is close to us, as something necessary for preserving the truth—and all the more so, given that we are philosophers. For though we love them both, piety bids us to honour the truth before our friends. (*EN* 1096a11–16)

Aristotle counters that Plato is wrong to assume that goodness is 'something universal, common to all good things, and single' (*EN* 1096a28). Rather, goodness is different in different cases.

To establish non-univocity, Aristotle's appeals to a variety of tests in his *Topics* where, again, his idiom is linguistic but his quarry is metaphysical. Consider the following sentences:

- Socrates is good.
- Communism is good.
- After a light meal, crème brûlée is good.
- Redoubling one's effort after failure is always good.

- Maria's singing is good, but Renata's is sublime.

Among the tests for non-univocity recommended in the *Topics* is a simple paraphrase test: if paraphrases yield distinct, non-interchangeable accounts, then the predicate is multivocal. So, for example, suitable paraphrases might be:

- Socrates is *a virtuous person*.
- Communism is *just social system*.
- After a light meal, crème brûlée is *tasty and satisfying*.
- Trying harder after one has failed is always *edifying*.
- Maria's singing *reaches a high artistic standard*, but Renata's *surpasses that standard by any measure*.

Since we cannot interchange these paraphrases—we cannot say, for instance, that crème brûlée is a just social system—*good* must be non-univocal across this range of applications. If that is correct, then Platonists are wrong to assume univocity in this case, since goodness exhibits complexity ignored by their assumption.

So far, then, Aristotle's appeals to homonymy or multivocity are primarily destructive, in the sense that they attempt to undermine a Platonic presumption regarded by Aristotle as unsustainable. Importantly, just as Aristotle sees a positive as well as a negative role for dialectic in philosophy, so he envisages in addition to its destructive applications a philosophically constructive role for homonymy. To appreciate his basic idea, it serves to reflect upon a continuum of positions in philosophical analysis ranging from pure Platonic univocity to disaggregated Wittgensteinean family resemblance. One might in the face of a successful challenge to Platonic univocity assume that, for instance, the various cases of goodness have nothing in common across all cases, so that good things form at best a motley kind, of the sort championed by Wittgensteineans

enamored of the metaphor of family resemblances: all good things belong to a kind only in the limited sense that they manifest a tapestry of partially overlapping properties, as every member of a single family is unmistakably a member of that family even though there is no one physical attribute shared by all of those family members.

Aristotle insists that there is a *tertium quid* between family resemblance and pure univocity: he identifies, and trumpets, a kind of *core-dependent homonymy* (also referred to in the literature, with varying degrees of accuracy, as *focal meaning* and *focal connexion*).[10] Core-dependent homonyms exhibit a kind of order in multiplicity: although shy of univocity, because homonymous, such concepts do not devolve into patchwork family resemblances either. To rely upon one of Aristotle's own favorite illustrations, consider:

- Socrates is healthy.
- Socrates' exercise regimen is healthy.
- Socrates' complexion is healthy.

Aristotle assumes that his readers will immediately appreciate two features of these three predications of *healthy*. First, they are non-univocal, since the second is paraphraseable roughly as *promotes health* and the third as *is indicative of health*, whereas the first means, rather, something more fundamental, like *is sound of body* or *is functioning well*. Hence, *healthy* is non-univocal. Second, even so, the last two predications rely upon the first for their elucidations: each appeals to health in its core sense in an asymmetrical way. That is, any account of each of the latter two predications *must* allude to the first, whereas an account of the first makes no reference to the second or third in its account. So, suggests Aristotle, *health* is not only a homonym, but a *core-dependent homonym*: while not univocal neither is it a case of rank multivocity.

Aristotle's illustration does succeed in showing that there is conceptual space between mere family resemblance and pure univocity. So, he is right that these are not exhaustive options. The interest in this sort of result resides in its exportability to richer, if more abstract philosophical concepts. Aristotle appeals to homonymy frequently, across a full range of philosophical concepts including *justice*, *causation*, *love*, *life*, *sameness*, *goodness*, and *body*. His most celebrated appeal to core-dependent homonymy comes in the case of a concept so highly abstract that it is difficult to gauge his success without extended metaphysical reflection. This is his appeal to the core-dependent homonymy of *being*, which has inspired both philosophical and scholarly controversy.[11] At one point, Aristotle denies that there could be a science of being, on the grounds that there is no single genus *being* under which all and only beings fall (*SE* 11 172a9–15). One motivation for his reasoning this way may be that he regards the notion of a genus as ineliminably taxonomical and contrastive, [12] so that it makes ready sense to speak of a genus of being only if one can equally well speak of a genus of non-being—just as among living beings one can speak of the animals and the non-animals, viz. the plant kingdom. Since there are no non-beings, there accordingly can be no genus of non-being, and so, ultimately, no genus of being either. Consequently, since each science studies one essential kind arrayed under a single genus, there can be no science of being either.

Subsequently, without expressly reversing his judgment about the existence of a science of being, Aristotle announces that there is nonetheless a science of being *qua* being (*Met*. iv 4), first philosophy, which takes as its subject matter beings insofar as they are beings and thus considers all and only those features pertaining to beings as such—to beings, that is, not insofar as they are mathematical or physical or human beings, but insofar as they are beings, full stop. Although the matter is disputed, his recognition of this science evidently turns crucially on his commitment to the core-dependent homonymy of being itself.[13]

Although the case is not as clear and uncontroversial as Aristotle's relatively easy appeal to *health* (which is why, after all, he selected it as an illustration), we are supposed to be able upon reflection to detect an analogous core-dependence in the following instances of *exists*:

- Socrates exists.
- Socrates' location exists.
- Socrates' weighing 73 kilos exists.
- Socrates' being morose today exists.

Of course, the last three items on this list are rather awkward locutions, but this is because they strive to make explicit that we can speak of dependent beings as existing if we wish to do so—but only because of their dependence upon the core instance of being, namely substance. (Here it is noteworthy that 'primary substance' is the conventional and not very happy rendering of Aristotle's *protê ousia* in Greek, which means, more literally, 'primary being').[14] According to this approach, we would not have Socrates' weighing anything at all or feeling any way today were it not for the prior fact of his existence. So, *exists* in the first instance serves as the core instance of being, in terms of which the others are to be explicated. If this is correct, then, implies Aristotle, *being* is a core-dependent homonym; further, a science of being becomes possible, even though there is no genus of being, since it is finally possible to study all beings insofar as they are related to the core instance of being, and then also to study that core instance, namely substance, insofar as it serves as the prime occasion of being.

6. Category Theory

In speaking of beings which depend upon substance for their existence, Aristotle implicitly appeals to a foundational philosophical commitment which appears early in his thought and remains stable throughout his

entire philosophical career: his theory of categories. In what is usually regarded as an early work, *The Categories*, Aristotle rather abruptly announces:

> Of things said without combination, each signifies either: (i) a substance (*ousia*); (ii) a quantity; (iii) a quality; (iv) a relative; (v) where; (vi) when; (vii) being in a position; (viii) having; (ix) acting upon; or (x) a being affected. (*Cat*. 1b25–27)

Aristotle does little to frame his theory of categories, offering no explicit derivation of it, nor even specifying overtly what his theory of categories categorizes. If librarians categorize books and botanists categorize plants, then what does the philosophical category theorist categorize?

Aristotle does not say explicitly, but his examples make reasonably clear that he means to categorize the basic kinds of beings there may be. If we again take some clues from linguistic data, without inferring that the ultimate objects of categorization are themselves linguistic, we can contrast things said "with combination":

- Man runs.

with things said 'without combination':

- Man
- Runs

'Man runs' is truth-evaluable, whereas neither 'man' nor 'runs' is. Aristotle says that things of this sort *signify* entities, evidently extra-linguistic entities, which are thus, correlatively, in the first case sufficiently complex to be what makes the sentence 'Man runs' true, that is a *man running*, and in the second, items below the level of truth-making, so, e.g., an entity *man*, taken by itself, and an action *running*, taken by itself. If that

is correct, the entities categorized by the categories are the sorts of basic beings that fall below the level of truth-makers, or facts. Such beings evidently contribute, so to speak, to the facticity of facts, just as, in their linguistic analogues, nouns and verbs, things said 'without combination', contribute to the truth-evaluability of simple assertions. The constituents of facts contribute to facts as the semantically relevant parts of a proposition contribute to its having the truth conditions it has. Thus, the items categorized in Aristotle's categories are the constituents of facts. If it is a fact *that Socrates is pale*, then the basic beings in view are *Socrates* and *being pale*. In Aristotle's terms, the first is a *substance* and the second is a *quality*.

Importantly, these beings may be *basic* without being absolutely *simple*. After all, Socrates is made up of all manner of parts—arms and legs, organs and bones, molecules and atoms, and so on down. As a useful linguistic analogue, we may consider *phonemes*, which are basic, relative to the morphemes of a linguistic theory, and yet also complex, since they are made up of simpler sound components, which are irrelevant from the linguist's point of view because of their lying beneath the level of semantic relevance.

The theory of categories in total recognizes ten sorts of extra-linguistic basic beings:

Category	Illustration
Substance	man, horse
Quality	white, grammatical
Quantity	two-feet long
Relative	double, slave
Place	in the market
Time	yesterday, tomorrow
Position	lying, sitting
Having	has shoes on
Acting Upon	cutting, burning
Being Affected	being cut, being burnt

Although he does not say so overtly in the *Categories*, Aristotle evidently presumes that these ten categories of being are both exhaustive and irreducible, so that while there are no other basic beings, it is not possible to eliminate any one of these categories in favor of another.

Both claims have come in for criticism, and each surely requires defense. [15] Aristotle offers neither conviction a defense in his *Categories*. Nor, indeed, does he offer any principled grounding for just these categories of being, a circumstance which has left him open to further criticism from later philosophers, including famously Kant who, after lauding Aristotle for coming up with the idea of category theory, proceeds to excoriate him for selecting his particular categories on no principled basis whatsoever. Kant alleges that Aristotle picked his categories of being just as he happened to stumble upon them in his reveries (*Critique of Pure Reason*, A81/B107). According to Kant, then, Aristotle's categories are *groundless*. Philosophers and scholars both before and after Kant have sought to provide the needed grounding, whereas Aristotle himself mainly tends to

justify the theory of categories by putting it to work in his various philosophical investigations.

We have already implicitly encountered in passing two of Aristotle's appeals to category theory: (i) in his approach to time, which he comes to treat as a non-substantial being; and (ii) in his commitment to the core-dependent homonymy of being, which introduces some rather more contentious considerations. These may be revisited briefly to illustrate how Aristotle thinks that his doctrine of categories provides philosophical guidance where it is most needed.

Thinking first of time and its various puzzles, or *aporiai*, we saw that Aristotle poses a simple question: does time exist? He answers this question in the affirmative, but only because in the end he treats it as a categorically circumscribed question. He claims that 'time is the measure of motion with respect to the before and after' (*Phys*. 219b1–2). By offering this definition, Aristotle is able to advance the judgment that time does exist, because it is an entity in the category of quantity: time is to motion or change as length is to a line. Time thus exists, but like all items in any non-substance category, it exists in a dependent sort of way. Just as if there were no lines there would be no length, so if there were no change there would be no time. Now, this feature of Aristotle's theory of time has occasioned both critical and favorable reactions.[16] In the present context, however, it is important only that it serves to demonstrate how Aristotle handles questions of existence: they are, at root, questions about category membership. A question as to whether, e.g., universals or places or relations exist, is ultimately, for Aristotle, also a question concerning their category of being, if any.

As time is a dependent entity in Aristotle's theory, so too are all entities in categories outside of substance. This helps explain why Aristotle thinks it appropriate to deploy his apparatus of core-dependent homonymy in the

case of *being*. If we ask whether qualities or quantities exist, Aristotle will answer in the affirmative, but then point out also that as dependent entities they do not exist in the independent manner of substances. Thus, even in the relatively rarified case of *being*, the theory of categories provides a reason for uncovering core-dependent homonymy. Since all other categories of being depend upon substance, it should be the case that an analysis of any one of them will ultimately make asymmetrical reference to substance. Aristotle contends in his *Categories*, relying on a distinction that tracks essential (*said-of*) and accidental (*in*) predication, that:

> All other things are either *said-of* primary substances, which are their subjects, or are *in* them as subjects. Hence, if there were no primary substances, it would be impossible for anything else to exist. (*Cat*. 2b5–6)

If this is so, then, Aristotle infers, all the non-substance categories rely upon substance as the core of their being. So, he concludes, being qualifies as a case of core-dependent homonymy.

Now, one may challenge Aristotle's contentions here, first by querying whether he has established the non-univocity of *being* before proceeding to argue for its core-dependence. Be that as it may, if we allow its non-univocity, then, according to Aristotle, the apparatus of the categories provides ample reason to conclude that *being* qualifies as a philosophically significant instance of core-dependent homonymy.

In this way, Aristotle's philosophy of being and substance, like much else in his philosophy, relies upon an antecedent commitment to his theory of categories. Indeed, the theory of categories spans his entire career and serves as a kind of scaffolding for much of his philosophical theorizing, ranging from metaphysics and philosophy of nature to psychology and value theory.

For this reason, questions regarding the ultimate tenability of Aristotle's doctrine of categories take on a special urgency for evaluating much of his philosophy.

For more detail on the theory of categories and its grounding, see the entry on Aristotle's Categories.

7. The Four Causal Account of Explanatory Adequacy

Equally central to Aristotle's thought is his *four-causal explanatory scheme*. Judged in terms of its influence, this doctrine is surely one of his most significant philosophical contributions. Like other philosophers, Aristotle expects the explanations he seeks in philosophy and science to meet certain criteria of adequacy. Unlike some other philosophers, however, he takes care to state his criteria for adequacy explicitly; then, having done so, he finds frequent fault with his predecessors for failing to meet its terms. He states his scheme in a methodological passage in the second book of his *Physics*:

> One way in which cause is spoken of is that out of which a thing comes to be and which persists, e.g. the bronze of the statue, the silver of the bowl, and the genera of which the bronze and the silver are species.

> In another way cause is spoken of as the form or the pattern, i.e. what is mentioned in the account (*logos*) belonging to the essence and its genera, e.g. the cause of an octave is a ratio of 2:1, or number more generally, as well as the parts mentioned in the account (*logos*).

Further, the primary source of the change and rest is spoken of as a cause, e.g. the man who deliberated is a cause, the father is the cause of the child, and generally the maker is the cause of what is made and what brings about change is a cause of what is changed.

Further, the end (*telos*) is spoken of as a cause. This is that for the sake of which (*hou heneka*) a thing is done, e.g. health is the cause of walking about. 'Why is he walking about?' We say: 'To be healthy'—and, having said that, we think we have indicated the cause.

(*Phys*. 194b23–35)

Although some of Aristotle's illustrations are not immediately pellucid, his approach to explanation is reasonably straightforward.

Aristotle's attitude towards explanation is best understood first by considering a simple example he proposes in *Physics* ii 3. A bronze statue admits of various different dimensions of explanation. If we were to confront a statue without first recognizing what it was, we would, thinks Aristotle, spontaneously ask a series of questions about it. We would wish to know *what it is*, *what it is made of*, *what brought it about*, and *what it is for*. In Aristotle's terms, in asking these questions we are seeking knowledge of the statue's four *causes* (*aitia*): the formal, material, efficient, and final. According to Aristotle, when we have identified these four causes, we have satisfied a reasonable demand for explanatory adequacy.

More fully, the four-causal account of explanatory adequacy requires an investigator to cite these four causes:

material	that from which something is generated and out of which it is made, e.g. the bronze of a statue.
formal	the structure which the matter realizes and in terms of which it comes to be something determinate, e.g., the shape of the president, in virtue of which this quantity of bronze is said to be a statue of a president.
efficient	the agent responsible for a quantity of matter's coming to be informed, e.g. the sculptor who shaped the quantity of bronze into its current shape, the shape of the president.
final	the purpose or goal of the compound of form and matter, e.g. the statue was created for the purpose of honoring the president.

THE FOUR CAUSES

In *Physics* ii 3, Aristotle makes twin claims about this four-causal schema: (i) that citing all four causes is *necessary* for adequacy in explanation; and (ii) that these four causes are *sufficient* for adequacy in explanation. Each of these claims requires some elaboration and also some qualification.

As for the necessity claim, Aristotle does not suppose that all phenomena admit of all four causes. Thus, for example, coincidences lack final causes, since they do not occur for the sake of anything; that is, after all, what makes them coincidences. If a debtor is on his way to the market to buy milk and she runs into her creditor, who is on his way to the same market to buy bread, then she may agree to pay the money owed immediately. Although resulting in a wanted outcome, their meeting was not for the sake of settling the debt; nor indeed was it for the sake of anything at all. It was a simple co-incidence. Hence, it lacks a final cause. Similarly, if we think that there are mathematical or geometrical abstractions, for instance a triangle existing as an object of thought independent of any material realization, then the triangle will trivially lack a material cause.[17] Still,

these significant exceptions aside, Aristotle expects the vast majority of explanations to conform to his four-causal schema. In non-exceptional cases, a failure to specify all four of causes, is, he maintains, a failure in explanatory adequacy.

The sufficiency claim is exceptionless, though it may yet be misleading if one pertinent issue is left unremarked. In providing his illustration of the material cause Aristotle first cites the bronze of a statue and the silver of a bowl, and then mentions also 'the genera of which the bronze and the silver are species' (*Phys*. 194b25–27). By this he means the types of metal to which silver and bronze belong, or more generally still, simply *metal*. That is, one might specify the material cause of a statue more or less proximately, by specifying the character of the matter more or less precisely. Hence, when he implies that citing all four causes is sufficient for explanation, Aristotle does not intend to suggest that a citation at any level of generality suffices. He means to insist rather that there is no fifth kind of cause, that his preferred four cases subsume all kinds of cause. He does not argue for this conclusion fully, though he does challenge his readers to identify a kind of cause which qualifies as a sort distinct from the four mentioned (*Phys*. 195a4–5).

So far, then, Aristotle's four causal schema has whatever intuitive plausibility his illustrations may afford it. He does not rest content there, however. Instead, he thinks he can argue forcefully for the four causes as real explanatory factors, that is, as features which must be cited not merely because they make for satisfying explanations, but because they are genuinely operative causal factors, the omission of which renders any putative explanation objectively incomplete and so inadequate.

It should be noted that Aristotle's arguments for the four causes taken individually all proceed against the backdrop of the general connection he forges between causal explanation and knowledge. Because he thinks that

the four *aitia* feature in answers to knowledge-seeking questions (*Phys.* 194b18; *A Po.* 71 b 9–11, 94 a 20), some scholars have come to understand them more as *becauses* than as *causes*—that is, as explanations rather than as causes narrowly construed.[18] Most such judgments reflect an antecedent commitment to one or another view of causation and explanation—that causation relates events rather than propositions; that explanations are inquiry-relative; that causation is extensional and explanation intensional; that explanations must adhere to some manner of nomic-deductive model, whereas causes need not; or that causes must be prior in time to their effects, while explanations, especially intentional explanations, may appeal to states of affairs posterior in time to the actions they explain.

Generally, Aristotle does not respect these sorts of commitments. Thus, to the extent that they are defensible, his approach to *aitia* may be regarded as blurring the canons of causation and explanation. It should certainly not, however, be ceded up front that Aristotle is guilty of any such conflation, or even that scholars who render his account of the four *aitia* in causal terms have failed to come to grips with developments in causal theory in the wake of Hume. Rather, because of the lack of uniformity in contemporary accounts of causation and explanation, and a persistent and justifiable tendency to regard causal explanations as foundational relative to other sorts of explanations, we may legitimately wonder whether Aristotle's conception of the four *aitia* is in any significant way discontinuous with later, Humean-inspired approaches, and then again, to the degree that it is, whether Aristotle's approach suffers for the comparison. Be that as it may, we will do well when considering Aristotle's defense of his four *aitia* to bear in mind that controversy surrounds how best to construe his knowledge-driven approach to causation and explanation relative to some later approaches.

For more on the four causes in general, see the entry on Aristotle on Causality.

8. Hylomorphism

Central to Aristotle's four-causal account of explanatory adequacy are the notions of *matter* (*hulê*) and form (*eidos* or *morphê*). Together, they constitute one of his most fundamental philosophical commitments, to *hylomorphism*:

- Hylomorphism $=_{df}$ ordinary objects are composites of matter and form.

The appeal in this definition to 'ordinary objects' requires reflection, but as a first approximation, it serves to rely on the sorts of examples Aristotle himself employs when motivating hylomorphism: statues and houses, horses and humans. In general, we may focus on artefacts and familiar living beings. Hylomorphism holds that no such object is metaphysically simple, but rather comprises two distinct metaphysical elements, one formal and one material.

Aristotle's hylomorphism was formulated originally to handle various puzzles about change. Among the *endoxa* confronting Aristotle in his *Physics* are some striking challenges to the coherence of the very notion of change, owing to Parmenides and Zeno. Aristotle's initial impulse in the face of such challenges, as we have seen, is to preserve the appearances (*phainomena*), to explain how change is possible. Key to Aristotle's response to the challenges bequeathed him is his insistence that all change involves at least two factors: something persisting and something gained or lost. Thus, when Socrates goes to the beach and comes away sun-tanned, something continues to exist, namely Socrates, even while something is lost, his pallor, and something else gained, his tan. This is a

change in the category of quality, whence the common locution 'qualitative change'. If he gains weight, then again something remains, Socrates, and something is gained, in this case a quantity of matter. Accordingly, in this instance we have not a qualitative but a quantitative change.

In general, argues Aristotle, in whatever category a change occurs, something is lost and something gained *within* that category, even while something else, a substance, remains in existence, as the subject of that change. Of course, substances can come into or go out of existence, in cases of generation or destruction; and these are changes in the category of substance. Evidently even in cases of change in this category, however, something persists. To take an example favourable to Aristotle, in the case of the generation of a statue, the bronze persists, but it comes to acquire a new form, a substantial rather than accidental form. In all cases, whether substantial or accidental, the two-factor analysis obtains: something remains the same and something is gained or lost.

In its most rudimentary formulation, hylomorphism simply labels each of the two factors: what remains is *matter* and what is gained is *form*. Aristotle's hylomorphism quickly becomes much more complex, however, as the notions of matter and form are pressed into philosophical service. Importantly, matter and form come to be paired with another fundamental distinction, that between *potentiality* and *actuality*. Again in the case of the generation of a statue, we may say that the bronze is *potentially* a statue, but that it is an *actual* statue when and only when it is informed with the form of a statue. Of course, before being made into a statue, the bronze was also in potentiality a fair number of other artefacts—a cannon, a steam-engine, or a goal on a football pitch. Still, it was not in potentiality butter or a beach ball. This shows that potentiality is not the same as possibility: to say that x is potentially F is to say that x already has actual features in virtue of which it might be made to be F by the imposition of a

F form upon it. So, given these various connections, it becomes possible to define form and matter generically as

- form $=_{df}$ that which makes some matter which is potentially F actually F
- matter $=_{df}$ that which persists and which is, for some range of Fs, potentially F

Of course, these definitions are circular, but that is not in itself a problem: actuality and potentiality are, for Aristotle, fundamental concepts which admit of explication and description but do not admit of reductive analyses.

Encapsulating Aristotle's discussions of change in *Physics* i 7 and 8, and putting the matter more crisply than he himself does, we have the following simple argument for matter and form: (1) a necessary condition of there being change is the existence of matter and form; (2) there is change; hence (3) there are matter and form. The second premise is a *phainomenon*; so, if that is accepted without further defense, only the first requires justification. The first premise is justified by the thought that since there is no generation *ex nihilo*, in every instance of change something persists while something else is gained or lost. In substantial generation or destruction, a substantial form is gained or lost; in mere accidental change, the form gained or lost is itself accidental. Since these two ways of changing exhaust the kinds of change there are, in *every* instance of change there are two factors present. These are matter and form.

For these reasons, Aristotle intends his hylomorphism to be much more than a simple explanatory heuristic. On the contrary, he maintains, matter and form are mind-independent features of the world and must, therefore, be mentioned in any full explanation of its workings.

9. Aristotelian Teleology

We may mainly pass over as uncontroversial the suggestion that there are efficient causes in favor of the most controversial and difficult of Aristotle four causes, the final cause.[19] We should note before doing so, however, that Aristotle's commitment to efficient causation does receive a defense in Aristotle's preferred terminology; he thus does more than many other philosophers who take it as given that causes of an efficient sort are operative. Partly by way of criticizing Plato's theory of Forms, which he regards as inadequate because of its inability to account for change and generation, Aristotle observes that nothing potential can bring itself into actuality without the agency of an actually operative efficient cause. Since what is potential is always in potentiality relative to some range of actualities, and nothing becomes actual of its own accord—no pile of bricks, for instance, spontaneously organizes itself into a house or a wall —an actually operative agent is required for every instance of change. This is the efficient cause. These sorts of considerations also incline Aristotle to speak of the priority of actuality over potentiality: potentialities are made actual by actualities, and indeed are always potentialities for some actuality or other. The operation of some actuality upon some potentiality is an instance of efficient causation.

That said, most of Aristotle's readers do not find themselves in need of a defense of the existence of efficient causation. By contrast, most think that Aristotle does need to provide a defense of final causation. It is natural and easy for us to recognize final causal activity in the products of human craft: computers and can-openers are devices dedicated to the execution of certain tasks, and both their formal and material features will be explained by appeal to their functions. Nor is it a mystery where artefacts obtain their functions: we give them their functions. The ends of artefacts are the results of the designing activities of intentional agents. Aristotle recognizes these kinds of final causation, but also, and more

problematically, envisages a much greater role for teleology in natural explanation: nature exhibits teleology without design. He thinks, for instance, that living organisms not only have parts which require teleological explanation—that, for instance, kidneys are *for* purifying the blood and teeth are *for* tearing and chewing food—but that whole organisms, human beings and other animals, also have final causes.

Crucially, Aristotle denies overtly that the causes operative in nature are intention-dependent. He thinks, that is, that organisms have final causes, but that they did not come to have them by dint of the designing activities of some intentional agent or other. He thus denies that a necessary condition of *x*'s having a final cause is *x*'s being designed.

Although he has been persistently criticized for his commitment to such natural ends, Aristotle is not susceptible to a fair number of the objections standardly made to his view. Indeed, it is evident that whatever the merits of the most penetrating of such criticisms, much of the contumely directed at Aristotle is stunningly illiterate.[20] To take but one of any number of mind-numbing examples, the famous American psychologist B. F. Skinner reveals that 'Aristotle argued that a falling body accelerated because it grew more jubilant as it found itself nearer its home' (1971, 6). To anyone who has actually read Aristotle, it is unsurprising that this ascription comes without an accompanying textual citation. For Aristotle, as Skinner would portray him, rocks are conscious beings having end states which they so delight in procuring that they accelerate themselves in exaltation as they grow ever closer to attaining them. There is no excuse for this sort of intellectual slovenliness, when already by the late-nineteenth century, the German scholar Zeller was able to say with perfect accuracy that 'The most important feature of the Aristotelian teleology is the fact that it is neither anthropocentric nor is it due to the actions of a creator existing outside the world or even of a mere arranger of the world, but is always thought of as immanent in nature' (1883, §48).

Indeed, it is hardly necessary to caricature Aristotle's teleological commitments in order to bring them into critical focus. In fact, Aristotle offers two sorts of defenses of non-intentional teleology in nature, the first of which is replete with difficulty. He claims in *Physics* ii 8:

> For these [viz. teeth and all other parts of natural beings] and all other natural things come about as they do either always or for the most part, whereas nothing which comes about due to chance or spontaneity comes about always or for the most part. … If, then, these are either the result of coincidence or for the sake of something, and they cannot be the result of coincidence or spontaneity, it follows that they must be for the sake of something. Moreover, even those making these sorts of claims [viz. that everything comes to be by necessity] will agree that such things are natural. Therefore, that for the sake of which is present among things which come to be and exist by nature. (*Phys*. 198b32–199a8)

The argument here, which has been variously formulated by scholars,[21] seems doubly problematic.

In this argument Aristotle seems to introduce as a *phainomenon* that nature exhibits regularity, so that the parts of nature come about in patterned and regular ways. Thus, for instance, humans tend to have teeth arranged in a predictable sort of way, with incisors in the front and molars in the back. He then seems to contend, as an exhaustive and exclusive disjunction, that things happen either by chance or for the sake of something, only to suggest, finally, that what is 'always or for the most part'—what happens in a patterned and predictable way—is not plausibly thought to be due to chance. Hence, he concludes, whatever happens always or for the most part must happen for the sake of something, and so must admit of a teleological cause. Thus, teeth show up always or for the

most part with incisors in the front and molars in the back; since this is a regular and predictable occurrence, it cannot be due to chance. Given that whatever is not due to chance has a final cause, teeth have a final cause.

If so much captures Aristotle's dominant argument for teleology, then his view is unmotivated. The argument is problematic in the first instance because it assumes an exhaustive and exclusive disjunction between what is by chance and what is for the sake of something. But there are obviously other possibilities. Hearts beat not in order to make noise, but they do so always and not by chance. Second, and this is perplexing if we have represented him correctly, Aristotle is himself aware of one sort of counterexample to this view and is indeed keen to point it out himself: although, he insists, bile is regularly and predictably yellow, its being yellow is neither due simply to chance nor for the sake of anything. Aristotle in fact mentions many such counterexamples (*Part. An.* 676b16–677b10, *Gen. An.* 778a29–b6). It seems to follow, then, short of ascribing a straight contradiction to him, either that he is not correctly represented as we have interpreted this argument or that he simply changed his mind about the grounds of teleology. Taking up the first alternative, one possibility is that Aristotle is not really trying to *argue* for teleology from the ground up in *Physics* ii 8, but is taking it as already established that there are teleological causes, and restricting himself to observing that many natural phenomena, namely those which occur always or for the most part, are good candidates for admitting of teleological explanation.

That would leave open the possibility of a broader sort of motivation for teleology, perhaps of the sort Aristotle offers elsewhere in the *Physics*, when speaking about the impulse to find non-intention-dependent teleological causes at work in nature:

> This is most obvious in the case of animals other than man: they make things using neither craft nor on the basis of inquiry nor by

deliberation. This is in fact a source of puzzlement for those who wonder whether it is by reason or by some other faculty that these creatures work—spiders, ants and the like. Advancing bit by bit in this same direction it becomes apparent that even in plants features conducive to an end occur—leaves, for example, grow in order to provide shade for the fruit. If then it is both by nature and for an end that the swallow makes its nest and the spider its web, and plants grow leaves for the sake of the fruit and send their roots down rather than up for the sake of nourishment, it is plain that this kind of cause is operative in things which come to be and are by nature. And since nature is twofold, as matter and as form, the form is the end, and since all other things are for sake of the end, the form must be the cause in the sense of that for the sake of which. (*Phys*. 199a20–32)

As Aristotle quite rightly observes in this passage, we find ourselves regularly and easily speaking in teleological terms when characterizing non-human animals and plants. It is consistent with our so speaking, of course, that all of our easy language in these contexts is lax and careless, because unwarrantedly anthropocentric. We might yet demand that all such language be assiduously reduced to some non-teleological idiom when we are being scientifically strict and empirically serious, though we would first need to survey the explanatory costs and benefits of our attempting to do so. Aristotle considers and rejects some views hostile to teleology in *Physics* ii 8 and *Generation and Corruption* i.[22]

10. Substance

Once Aristotle has his four-causal explanatory schema fully on the scene, he relies upon it in virtually all of his most advanced philosophical investigation. As he deploys it in various frameworks, we find him augmenting and refining the schema even as he applies it, sometimes with

surprising results. One important question concerns how his hylomorphism intersects with the theory of substance advanced in the context of his theory of categories.

As we have seen, Aristotle insists upon the primacy of primary substance in his *Categories*. According to that work, however, star instances of primary substance are familiar living beings like Socrates or an individual horse (*Cat*. 2a11014). Yet with the advent of hylomorphism, these primary substances are revealed to be metaphysical complexes: Socrates is a compound of matter and form. So, now we have not one but three potential candidates for primary substance: form, matter, and the compound of matter and form. The question thus arises: which among them is the primary substance? Is it the matter, the form, or the compound? The compound corresponds to a basic object of experience and seems to be a basic subject of predication: we say that Socrates lives in Athens, not that his matter lives in Athens. Still, matter underlies the compound and in this way seems a more basic subject than the compound, at least in the sense that it can exist before and after it does. On the other hand, the matter is nothing definite at all until enformed; so, perhaps form, as determining what the compound is, has the best claim on substantiality.

In the middle books of his *Metaphysics*, which contain some of his most complex and engaging investigations into basic being, Aristotle settles on *form* (*Met*. vii 17). A question thus arises as to how form satisfies Aristotle's final criteria for substantiality. He expects a substance to be, as he says, some particular thing (*tode ti*), but also to be something knowable, some essence or other. These criteria seem to pull in different directions, the first in favor of particular substances, as the primary substances of the *Categories* had been particulars, and the second in favor of universals as substances, because they alone are knowable. In the lively controversy surrounding these matters, many scholars have concluded that

Aristotle adopts a third way forward: form is both knowable *and* particular. This matter, however, remains very acutely disputed.[23]

Very briefly, and not engaging these controversies, it becomes clear that Aristotle prefers form in virtue of its role in generation and diachronic persistence. When a statue is generated, or when a new animal comes into being, something persists, namely the matter, which comes to realize the substantial form in question. Even so, insists Aristotle, the matter does not by itself provide the identity conditions for the new substance. First, as we have seen, the matter is merely potentially some *F* until such time as it is made actually *F* by the presence of an *F* form. Further, the matter can be replenished, and *is* replenished in the case of all organisms, and so seems to be form-dependent for its own diachronic identity conditions. For these reasons, Aristotle thinks of the form as prior to the matter, and thus more fundamental than the matter. This sort of matter, the form-dependent matter, Aristotle regards as *proximate matter* (*Met*. 1038b6, 1042b10), thus extending the notion of matter beyond its original role as metaphysical substrate.

Further, in *Metaphysics* vii 17 Aristotle offers a suggestive argument to the effect that matter alone cannot be substance. Let the various bits of matter belonging to Socrates be labeled as $a, b, c, \ldots, n$. Consistent with the non-existence of Socrates is the existence of $a, b, c, \ldots, n$, since these elements exist when they are spread from here to Alpha Centauri, but if that happens, of course, Socrates no longer exists. Heading in the other direction, Socrates can exist without just these elements, since he may exist when some one of $a, b, c, \ldots, n$ is replaced or goes out of existence. So, in addition to his material elements, insists Aristotle, Socrates is also something else, something more (*heteron ti*; *Met*. 1041b19–20). This something more is *form*, which is 'not an element…but a primary cause of a thing's being what it is' (*Met*. 1041b28–30). The cause of a thing's being

the actual thing it is, as we have seen, is form. Hence, concludes Aristotle, as the source of being and unity, form is substance.

Even if this much is granted—and to repeat, much of what has just been said is unavoidably controversial—many questions remain. For example, is form best understood as universal or particular? However that issue is to be resolved, what is the relation of form to the compound and to matter? If form is substance, then what is the fate of these other two candidates? Are they also substances, if to a lesser degree? It seems odd to conclude that they are nothing at all, or that the compound in particular is nothing in actuality; yet it is difficult to contend that they might belong to some category other than substance.

For an approach to some of these questions, see the entry on Aristotle's Metaphysics.

11. Living Beings

However these and like issues are to be resolved, given the primacy of form as substance, it is unsurprising to find Aristotle identifying the soul, which he introduces as a principle or source (*archê*) of all life, as the form of a living compound. For Aristotle, in fact, all living things, and not only human beings, have souls: 'what is ensouled is distinguished from what is unensouled by living' (*DA* 431a20–22; cf. *DA* 412a13, 423a20–6; *De Part. An.* 687a24–690a10; *Met.* 1075a16–25). It is appropriate, then, to treat all ensouled bodies in hylomorphic terms:

> The soul is the cause and source of the living body. But *cause* and *source* are meant in many ways [or are homonymous]. Similarly, the soul is a cause in accordance with the ways delineated, which are three: it is (i) the cause as the source of motion [=the efficient cause], (ii) that for the sake of which [=the final cause], and (iii) as

the substance of ensouled bodies. That it is a cause as substance is clear, for substance is the cause of being for all things, and for living things, being is life, and the soul is also the cause and source of life. (*DA* 415b8–14; cf. *PN* 467b12–25, *Phys*. 255a56–10)

So, the soul and body are simply special cases of form and matter:

soul : body :: form : matter :: actuality : potentiality

Further, the soul, as the end of the compound organism, is also the final cause of the body. Minimally, this is to be understood as the view that any given body is the body that it is because it is organized around a function which serves to unify the entire organism. In this sense, the body's unity derives from the fact it has a single end, or single life directionality, a state of affairs that Aristotle captures by characterizing the body as the sort of matter which is *organic* (*organikon*; *DA* 412a28). By this he means that the body serves as a tool for implementing the characteristic life activities of the kind to which the organism belongs (*organon = tool* in Greek). Taking all this together, Aristotle offers the view that the soul is the 'first actuality of a natural organic body' (*DA* 412b5–6), that it is a 'substance as form of a natural body which has life in potentiality' (*DA* 412a20–1) and, again, that it 'is a first actuality of a natural body which has life in potentiality' (*DA* 412a27–8).

Aristotle contends that his hylomorphism provides an attractive middle way between what he sees as the mirroring excesses of his predecessors. In one direction, he means to reject Presocratic kinds of materialism; in the other, he opposes Platonic dualism. He gives the Presocratics credit for identifying the material causes of life, but then faults them for failing to grasp its formal cause. By contrast, Plato earns praise for grasping the formal cause of life; unfortunately, he then proceeds to neglect the material cause, and comes to believe that the soul can exist without its

material basis. Hylomorphism, in Aristotle's view, captures what is right in both camps while eschewing the unwarranted mono-dimensionality of each. In his view, to account for living organisms, one must attend to both matter and form.

Aristotle deploys hylomorphic analyses not only to the whole organism, but to the individual faculties of the soul as well. Perception involves the reception of sensible forms without matter, and thinking, by analogy, consists in the mind's being enformed by intelligible forms. With each of these extensions, Aristotle both expands and taxes his basic hylomorphism, sometimes straining its basic framework almost beyond recognition.

For more detail on Aristotle's hylomorphism in psychological explanation, see the entry on Aristotle's Psychology.

12. Happiness and Political Association

Aristotle's basic teleological framework extends to his ethical and political theories, which he regards as complementing one another. He takes it as given that most people wish to lead good lives; the question then becomes what the best life for human beings consists in. Because he believes that the best life for a human being is not a matter of subjective preference, he also believes that people can (and, sadly, often do) choose to lead sub-optimal lives. In order to avoid such unhappy eventualities, Aristotle recommends reflection on the criteria any successful candidate for the best life must satisfy. He proceeds to propose one kind of life as meeting those criteria uniquely and therefore promotes it as the superior form of human life. This is a life lived in accordance with reason.

When stating the general criteria for the final good for human beings, Aristotle invites his readers to review them (*EN* 1094a22–27). This is

advisable, since much of the work of sorting through candidate lives is in fact accomplished during the higher-order task of determining the criteria appropriate to this task. Once these are set, it becomes relatively straightforward for Aristotle to dismiss some contenders, including for instance the life of pleasure.

According to the criteria advanced, the final good for human beings must: (i) be pursued for its own sake (*EN* 1094a1); (ii) be such that we wish for other things for its sake (*EN* 1094a19); (iii) be such that we do not wish for it on account of other things (*EN* 1094a21); (iv) be complete (*teleion*), in the sense that it is always choiceworthy and always chosen for itself (*EN* 1097a26–33); and finally (v) be self-sufficient (*autarkês*), in the sense that its presence suffices to make a life lacking in nothing (*EN* 1097b6–16). Plainly some candidates for the best life fall down in the face of these criteria. According to Aristotle, neither the life of pleasure nor the life of honour satisfies them all.

What does satisfy them all is happiness *eudaimonia*. Scholars in fact dispute whether *eudaimonia* is best rendered as 'happiness' or 'flourishing' or 'living well' or simply transliterated and left an untranslated technical term.[24] If we have already determined that *happiness* is some sort of subjective state, perhaps simple desire fulfillment, then 'happiness' will indeed be an inappropriate translation: *eudaimonia* is achieved, according to Aristotle, by fully realizing our natures, by actualizing to the highest degree our human capacities, and neither our nature nor our endowment of human capacities is a matter of choice for us. Still, as Aristotle frankly acknowledges, people will consent without hesitation to the suggestion that happiness is our best good—even while differing materially about how they understand what happiness is. So, while seeming to agree, people in fact disagree about the human good. Consequently, it is necessary to reflect on the nature of happiness (*eudaimonia*):

But perhaps saying that the highest good is happiness (*eudaimonia*) will appear to be a platitude and what is wanted is a much clearer expression of what this is. Perhaps this would come about if the function (*ergon*) of a human being were identified. For just as the good, and doing well, for a flute player, a sculptor, and every sort of craftsman—and in general, for whatever has a function and a characteristic action—seems to depend upon function, so the same seems true for a human being, if indeed a human being has a function. Or do the carpenter and cobbler have their functions, while a human being has none and is rather naturally without a function (*argon*)? Or rather, just as there seems to be some particular function for the eye and the hand and in general for each of the parts of a human being, should one in the same way posit a particular function for the human being in addition to all these? Whatever might this be? For living is common even to plants, whereas something characteristic (*idion*) is wanted; so, one should set aside the life of nutrition and growth. Following that would be some sort of life of perception, yet this is also common, to the horse and the bull and to every animal. What remains, therefore, is a life of action belonging to the kind of soul that has reason. (*EN* 1097b22–1098a4)

In determining what *eudaimonia* consists in, Aristotle makes a crucial appeal to the human function (*ergon*), and thus to his overarching teleological framework.

He thinks that he can identify the human function in terms of reason, which then provides ample grounds for characterizing the happy life as involving centrally the exercise of reason, whether practical or theoretical. Happiness turns out to be an activity of the rational soul, conducted in accordance with virtue or excellence, or, in what comes to the same thing, in rational activity executed excellently (*EN* 1098a161–17). It bears noting

in this regard that Aristotle's word for virtue, *aretê*, is broader than the dominant sense of the English word 'virtue', since it comprises all manner of excellences, thus including but extending beyond the moral virtues. Thus when he says that happiness consists in an activity in 'accordance with virtue' (*kat' aretên*; *EN* 1098a18), Aristotle means that it is a kind of excellent activity, and not merely morally virtuous activity.

The suggestion that only *excellently* executed or *virtuously* performed rational activity constitutes human happiness provides the impetus for Aristotle's virtue ethics. Strikingly, first, he insists that the good life is a life of activity; no *state* suffices, since we are commended and praised for living good lives, and we are rightly commended or praised only for things we do (*EN* 1105b20–1106a13). Further, given that we must not only act, but act excellently or virtuously, it falls to the ethical theorist to determine what virtue or excellence consists in with respect to the individual human virtues, including, for instance, courage and practical intelligence. This is why so much of Aristotle's ethical writing is given over to an investigation of virtue, both in general and in particular, and extending to both practical and theoretical forms.

For more on Aristotle's virtue-based ethics, see the entry on Aristotle's Ethics.

Aristotle concludes his discussion of human happiness in his *Nicomachean Ethics* by introducing political theory as a continuation and completion of ethical theory. Ethical theory characterizes the best form of human life; political theory characterizes the forms of social organization best suited to its realization (*EN* 1181b12–23).

The basic political unit for Aristotle is the *polis*, which is both a *state* in the sense of being an authority-wielding monopoly and a *civil society* in the sense of being a series of organized communities with varying degrees

of converging interest. Aristotle's political theory is markedly unlike some later, liberal theories, in that he does not think that the *polis* requires justification as a body threatening to infringe on antecedently existing human rights. Rather, he advances a form of political naturalism which treats human beings as by nature political animals, not only in the weak sense of being gregariously disposed, nor even in the sense of their merely benefiting from mutual commercial exchange, but in the strong sense of their flourishing as human beings at all *only* within the framework of an organized *polis*. The *polis* 'comes into being for the sake of living, but it remains in existence for the sake of living well' (*Pol*. 1252b29–30; cf. 1253a31–37).

The *polis* is thus to be judged against the goal of promoting human happiness. A superior form of political organization enhances human life; an inferior form hampers and hinders it. One major question pursued in Aristotle's *Politics* is thus structured by just this question: what sort of political arrangement best meets the goal of developing and augmenting human flourishing? Aristotle considers a fair number of differing forms of political organization, and sets most aside as inimical to the goal human happiness. For example, given his overarching framework, he has no difficulty rejecting contractarianism on the grounds that it treats as merely instrumental those forms of political activity which are in fact partially constitutive of human flourishing (*Pol*. iii 9).

In thinking about the possible kinds of political organization, Aristotle relies on the structural observations that rulers may be one, few, or many, and that their forms of rule may be legitimate or illegitimate, as measured against the goal of promoting human flourishing (*Pol*. 1279a26–31). Taken together, these factors yield six possible forms of government, three correct and three deviant:

	Correct	**Deviant**
One Ruler	Kingship	Tyranny
Few Rulers	Aristocracy	Oligarchy
Many Rulers	Polity	Democracy

The correct are differentiated from the deviant by their relative abilities to realize the basic function of the *polis*: living well. Given that we prize human happiness, we should, insists Aristotle, prefer forms of political association best suited to this goal.

Necessary to the end of enhancing human flourishing, maintains Aristotle, is the maintenance of a suitable level of distributive justice. Accordingly, he arrives at his classification of better and worse governments partly by considerations of distributive justice. He contends, in a manner directly analogous to his attitude towards *eudaimonia*, that everyone will find it easy to agree to the proposition that we should prefer a just state to an unjust state, and even to the formal proposal that the distribution of justice requires treating equal claims similarly and unequal claims dissimilarly. Still, here too people will differ about what constitutes an equal or an unequal claim or, more generally, an equal or an unequal person. A democrat will presume that all citizens are equal, whereas an aristocrat will maintain that the best citizens are, quite obviously, superior to the inferior. Accordingly, the democrat will expect the formal constraint of justice to yield equal distribution to all, whereas the aristocrat will take for granted that the best citizens are entitled to more than the worst.

When sorting through these claims, Aristotle relies upon his own account of distributive justice, as advanced in *Nicomachean Ethics* v 3. That account is deeply meritocratic. He accordingly disparages oligarchs, who suppose that justice requires preferential claims for the rich, but also democrats, who contend that the state must boost liberty across all citizens

irrespective of merit. The best *polis* has neither function: its goal is to enhance human flourishing, an end to which liberty is at best instrumental, and not something to be pursued for its own sake.

Still, we should also proceed with a sober eye on what is in fact possible for human beings, given our deep and abiding acquisitional propensities. Given these tendencies, it turns out that although deviant, democracy may yet play a central role in the sort of mixed constitution which emerges as the best form of political organization available to us. Inferior though it is to polity (that is, rule by the many serving the goal of human flourishing), and especially to aristocracy (government by the best humans, the *aristoi*, also dedicated to the goal of human flourishing), democracy, as the best amongst the deviant forms of government, may also be the most we can realistically hope to achieve.

For an in-depth discussion of Aristotle's political theory, including his political naturalism, see the entry on Aristotle's Politics.

13. Rhetoric and the Arts

Aristotle regards rhetoric and the arts as belonging to the productive sciences. As a family, these differ from the practical sciences of ethics and politics, which concern human conduct, and from the theoretical sciences, which aim at truth for its own sake. Because they are concerned with the creation of human products broadly conceived, the productive sciences include activities with obvious, artefactual products like ships and buildings, but also agriculture and medicine, and even, more nebulously, rhetoric, which aims at the production of persuasive speech (*Rhet*. 1355b26; cf. *Top*. 149b5), and tragedy, which aims at the production of edifying drama (*Poet*. 1448b16–17). If we bear in mind that Aristotle approaches all these activities within the broader context of his teleological explanatory framework, then at least some of the highly

polemicized interpretative difficulties which have grown up around his works in this area, particularly the *Poetics*, may be sharply delimited.

One such controversy centers on the question of whether Aristotle's *Rhetoric* and *Poetics* are primarily descriptive or prescriptive works.[25] To the degree that they are indeed prescriptive, one may wonder whether Aristotle has presumed in these treatises to dictate to figures of the stature of Sophocles and Euripides how best to pursue their crafts. To some extent —but only to some extent—it may seem that he does. There are, at any rate, clearly prescriptive elements in both these texts. Still, he does not arrive at these recommendations *a priori*. Rather, it is plain that Aristotle has collected the best works of forensic speech and tragedy available to him, and has studied them to discern their more and less successful features. In proceeding in this way, he aims to capture and codify what is best in both rhetorical practice and tragedy, in each case relative to its appropriate productive goal.

The general goal of rhetoric is clear. Rhetoric, says Aristotle, 'is the power to see, in each case, the possible ways to persuade' (*Rhet*. 1355b26). Different contexts, however, require different techniques. Thus, suggests Aristotle, speakers will usually find themselves in one of three contexts where persuasion is paramount: deliberative (*Rhet*. i 4–8), epideictic (*Rhet*. i 9), and judicial (*Rhet*. i 10–14). In each of these contexts, speakers will have at their disposal three main avenues of persuasion: the character of the speaker, the emotional constitution of the audience, and the general argument (*logos*) of the speech itself (*Rhet*. i 3). Rhetoric thus examines techniques of persuasion pursuant to each of these areas.

When discussing these techniques, Aristotle draws heavily upon topics treated in his logical, ethical, and psychological writings. In this way, the *Rhetoric* illuminates Aristotle's writings in these comparatively theoretical areas by developing in concrete ways topics treated more abstractly

elsewhere. For example, because a successful persuasive speech proceeds alert to the emotional state of the audience on the occasion of its delivery, Aristotle's *Rhetoric* contains some of his most nuanced and specific treatments of the emotions. Heading in another direction, a close reading of the *Rhetoric* reveals that Aristotle treats the art of persuasion as closely akin to dialectic (see §4.3 above). Like dialectic, rhetoric trades in techniques that are not scientific in the strict sense (see §4.2 above), and though its goal is persuasion, it reaches its end best if it recognizes that people naturally find proofs and well-turned arguments persuasive (*Rhet*. 1354a1, 1356a25, 1356a30). Accordingly, rhetoric, again like dialectic, begins with credible opinions (*endoxa*), though mainly of the popular variety rather than those endorsed most readily by the wise (*Top*. 100a29–35; 104a8–20; *Rhet*. 1356b34). Finally, rhetoric proceeds from such opinions to conclusions which the audience will understand to follow by cogent patterns of inference (*Rhet*. 1354a12–18, 1355a5–21). For this reason, too, the rhetorician will do well understand the patterns of human reasoning.

For more on Aristotle's rhetoric, see the entry on Aristotle's Rhetoric.

By highlighting and refining techniques for successful speech, the *Rhetoric* is plainly prescriptive—but only relative to the goal of persuasion. It does not, however, select its own goal or in any way dictate the end of persuasive speech: rather, the end of rhetoric is given by the nature of the craft itself. In this sense, the *Rhetoric* is like both the *Nicomachean Ethics* and the *Politics* in bearing the stamp of Aristotle's broad and encompassing teleology.

The same holds true of the *Poetics*, but in this case the end is not easily or uncontroversially articulated. It is often assumed that the goal of tragedy is *catharsis*—the purification or purgation of the emotions aroused in a tragic performance. Despite its prevalence, as an interpretation of what Aristotle

actually says in the *Poetics* this understanding is underdetermined at best. When defining tragedy in a general way, Aristotle claims:

> Tragedy, then, is an imitation of an action that is serious and complete, and which has some greatness about it. It imitates in words with pleasant accompaniments, each type belonging separately to the different parts of the work. It imitates people performing actions and does not rely on narration. It achieves, through pity and fear, the catharsis of these sorts of feelings. (*Poet.* 1449b21–29)

Although he has been represented in countless works of scholarship as contending that tragedy *is for the sake of catharsis*, Aristotle is in fact far more circumspect. While he does contend that tragedy will effect or accomplish catharsis, in so speaking he does not use language which clearly implies that catharsis is in itself the function of tragedy. Although a good blender will achieve a blade speed of 36,000 rotations per minute, this is not its function; rather, it achieves this speed in service of its function, namely blending. Similarly, then, on one approach, tragedy achieves catharsis, though not because it is its function to do so. This remains so, even if it is integral to realizing its function that tragedy achieve catharsis—as it is equally integral that it makes us of imitation (*mimêsis*), and does so by using words along with pleasant accompaniments (namely, rhythm, harmony, and song; *Poet.* 1447b27).

Unfortunately, Aristotle is not completely forthcoming on the question of the function of tragedy. One clue towards his attitude comes from a passage in which he differentiates tragedy from historical writing:

> The poet and the historian differ not in that one writes in meter and the other not; for one could put the writings of Herodotus into verse and they would be none the less history, with or without

meter. The difference resides in this: the one speaks of what has happened, and the other of what might be. Accordingly, poetry is more philosophical and more momentous than history. The poet speaks more of the universal, while the historian speaks of particulars. It is universal that when certain things turn out a certain way someone will in all likelihood or of necessity act or speak in a certain way—which is what the poet, though attaching particular names to the situation, strives for. (*Poet*. 1451a38–1451b10)

In characterizing poetry as more philosophical, universal, and momentous than history, Aristotle praises poets for their ability to assay deep features of human character, to dissect the ways in which human fortune engages and tests character, and to display how human foibles may be amplified in uncommon circumstances. We do not, however, reflect on character primarily for entertainment value. Rather, and in general, Aristotle thinks of the goal of tragedy in broadly intellectualist terms: the function of tragedy is 'learning, that is, figuring out what each thing is' (*Poet*. 1448b16–17). In Aristotle's view, tragedy teaches us about ourselves.

That said, catharsis is undoubtedly a key concept in Aristotle's *Poetics*, one which, along with imitation (*mimêsis*), has generated enormous controversy.[26] These controversies center around three poles of interpretation: the *subject* of catharsis, the *matter* of the catharsis, and the *nature* of catharsis. To illustrate what is meant: on a naïve understanding of catharsis—which may be correct despite its naïveté—the *audience* (the subject) undergoes catharsis by having the *emotions* (the matter) of pity and fear it experiences *purged* (the nature). By varying just these three possibilities, scholars have produced a variety of interpretations—that it is the actors or even the plot of the tragedy which are the subjects of catharsis, that the purification is cognitive or structural rather than emotional, and that catharsis is purification rather than purgation. On this

last contrast, just as we might purify blood by filtering it, rather than purging the body of blood by letting it, so we might refine our emotions, by cleansing them of their more unhealthy elements, rather than ridding ourselves of the emotions by purging them altogether. The difference is considerable, since on one view the emotions are regarded as in themselves destructive and so to be purged, while on the other, the emotions may be perfectly healthy, even though, like other psychological states, they may be improved by refinement. The immediate context of the *Poetics* does not by itself settle these disputes conclusively.

Aristotle says comparatively more about the second main concept of the *Poetics*, imitation (*mimêsis*). Although less controversial than catharsis, Aristotle's conception of *mimêsis* has also been debated.[27] Aristotle thinks that imitation is a deeply ingrained human proclivity. Like political association, he contends, *mimêsis* is *natural*. We engage in imitation from an early age, already in language learning by aping competent speakers as we learn, and then also later, in the acquisition of character by treating others as role models. In both these ways, we imitate because we learn and grow by imitation, and for humans, learning is both natural and a delight (*Poet*. 1148b4–24). This same tendency, in more sophisticated and complex ways, leads us into the practice of drama. As we engage in more advanced forms of *mimêsis*, imitation gives way to *representation* and *depiction*, where we need not be regarded as attempting to *copy* anyone or anything in any narrow sense of the term. For tragedy does not set out merely to copy what is the case, but rather, as we have seen in Aristotle's differentiation of tragedy from history, to speak of what might be, to engage universal themes in a philosophical manner, and to enlighten an audience by their depiction. So, although *mimêsis* is at root simple imitation, as it comes to serve the goals of tragedy, it grows more sophisticated and powerful, especially in the hands of those poets able to deploy it to good effect.

14. Aristotle's Legacy

Aristotle's influence is difficult to overestimate. After his death, his school, the Lyceum, carried on for some period of time, though precisely how long is unclear. In the century immediately after his death, Aristotle's works seem to have fallen out of circulation; they reappear in the first century B.C.E., after which time they began to be disseminated, at first narrowly, but then much more broadly. They eventually came to form the backbone of some seven centuries of philosophy, in the form of the commentary tradition, much of it original philosophy carried on in a broadly Aristotelian framework. They also played a very significant, if subordinate role, in the Neoplatonic philosophy of Plotinus and Porphyry. Thereafter, from the sixth through the twelfth centuries, although the bulk of Aristotle's writings were lost to the West, they received extensive consideration in Byzantine Philosophy, and in Arabic Philosophy, where Aristotle was so prominent that be became known simply as The First Teacher (see the entry on the influence of Arabic and Islamic philosophy on the Latin West). In this tradition, the notably rigorous and illuminating commentaries of Avicenna and Averroes interpreted and developed Aristotle's views in striking ways. These commentaries in turn proved exceedingly influential in the earliest reception of the Aristotelian corpus into the Latin West in the twelfth century.

Among Aristotle's greatest exponents during the early period of his reintroduction to the West, Albertus Magnus, and above all his student Thomas Aquinas, sought to reconcile Aristotle's philosophy with Christian thought. Some Aristotelians disdain Aquinas as bastardizing Aristotle, while some Christians disown Aquinas as pandering to pagan philosophy. Many others in both camps take a much more positive view, seeing Thomism as a brilliant synthesis of two towering traditions; arguably, the incisive commentaries written by Aquinas towards the end of his life aim not so much at synthesis as straightforward exegesis and exposition, and in

these respects they have few equals in any period of philosophy. Partly due to the attention of Aquinas, but for many other reasons as well, Aristotelian philosophy set the framework for the Christian philosophy of the twelfth through the sixteenth centuries, though, of course, that rich period contains a broad range of philosophical activity, some more and some less in sympathy with Aristotelian themes. To see the extent of Aristotle's influence, however, it is necessary only to recall that the two concepts forming the so-called *binarium famosissimum* ("the most famous pair") of that period, namely universal hylomorphism and the doctrine of the plurality of forms, found their first formulations in Aristotle's texts.

Interest in Aristotle continued unabated throughout the renaissance in the form of Renaissance Aristotelianism. The dominant figures of this period overlap with the last flowerings of Medieval Aristotelian Scholasticism, which reached a rich and highly influential close in the figure of Suárez, whose life in turn overlaps with Descartes. From the end of late Scholasticism, the study of Aristotle has undergone various periods of relative neglect and intense interest, but has been carried forward uninterrupted down to the present day.

Today, philosophers of various stripes continue to look to Aristotle for guidance and inspiration in many different areas, ranging from the philosophy of mind to theories of the infinite, though perhaps Aristotle's influence is seen most overtly and avowedly in the resurgence of virtue ethics which began in the last half of the twentieth century. It seems safe at this stage to predict that Aristotle's stature is unlikely to diminish in the new millennium. If it is any indication of the direction of things to come, a quick search of the present Encyclopedia turns up more citations to 'Aristotle' and 'Aristotelianism' than to any other philosopher or philosophical movement. Only Plato comes close.

Bibliography

Medieval Philosophy

First published Tue Aug 3, 2004; substantive revision Tue Mar 15, 2016

Medieval philosophy is conventionally construed as the philosophy of Western Europe between the decline of classical pagan culture and the Renaissance. Such a broad topic cannot be covered in detail in a single article, and fortunately there is no need to do so, since other articles in this *Encyclopedia* treat individual medieval philosophers and topics. The present article will confine itself to articulating some of the overall contours of medieval philosophy. The reader should refer to the items listed under Related Entries below for more detailed information on narrower subjects.

- 1. The Geographical and Chronological Boundaries of Medieval Philosophy
- 2. The Main Ingredients of Medieval Philosophy
- 3. The Availability of Greek Texts
- 4. From the Patristic Period to the Mid-Twelfth Century
    - 4.1 Augustine
    - 4.2 Boethius
    - 4.3 The Carolingian Period
    - 4.4 Anselm of Canterbury
    - 4.5 Peter Abelard
    - 4.6 General Characteristics of This Early Period
- 5. The Twelfth Century and the Rise of Universities
    - 5.1 New Translations
    - 5.2 New Forms of Education
- 6. The Thirteenth Century and Later
- 7. Some Main Topics in Medieval Philosophy
- Bibliography
- Academic Tools

1. The Geographical and Chronological Boundaries of Medieval Philosophy

'Medieval philosophy' refers to philosophy in Western Europe during the "medieval" period, the so called "Middle Ages." The notion of a "Middle Age" (or plural "Middle Ages") was introduced in the fifteenth century for the period between the decline of classical pagan culture in Western Europe and what was taken to be its rediscovery during the Renaissance. The first known documented use of the expression (in the form '*media tempestas*') is from 1469 (Robinson [1984], p. 748).[1]

The originators of the notion of the Middle Ages were thinking primarily of the so called "Latin West," the area, roughly speaking, of Roman Catholicism. While it is true that this region was to some extent a unit, culturally separate from its neighbors, it is also true that medieval philosophy was decisively influenced by ideas from the Greek East, from the Jewish philosophical tradition, and from Islam. If one takes medieval philosophy to include the Patristic period, as the present author prefers to do, then the area must be expanded to include, at least during the early centuries, Greek-speaking eastern Europe, as well as North Africa and parts of Asia Minor.

The chronological limits of medieval philosophy are likewise imprecise. Many histories of medieval philosophy (like many syllabi for courses on the subject) begin with St. Augustine (354–430), though some include second- and third-century Christian thinkers (see Marenbon [2007], p. 1), whereas Pasnau ([2010], p. 1) speaks of a more recent "consensus on when and where to place the beginnings of medieval philosophy,

understood as a project of independent philosophical inquiry: it begins in Baghdad, in the middle of the eighth century, and in France, in the itinerant court of Charlemagne, in the last quarter of the eighth century." At the other end of the period, things are even more imprecise. Robinson ([1984], pp. 749–50) amusingly summarizes the situation:[2]

> Scholars have advocated many different termini for our period, and there seems to be little agreement and indeed little basis for reasoned argument on these points. The Middle Ages begin, we are told, with the death of Theodosius in 395, or with the settlement of Germanic tribes in the Roman Empire, or with the sack of Rome in 410, or with the fall of the Western Roman Empire (usually dated C.E. 476), or even as late as the Moslem occupation of the Mediterranean. It ends … with the fall of Constantinople, or with the invention of printing, or with the discovery of America, or with the beginning of the Italian wars (1494), or with the Lutheran Reformation (1517), or with the election of Charles V (1519). Several reference works I have consulted simply assert that the Middle Ages ended in 1500, presumably on New Year's Eve. Yet another terminus often given for the Middle Ages is the so-called "Revival of Learning," that marvelous era when Humanist scholars "discovered" classical texts and restored them to mankind after the long Gothic night. Medievalists must always smile a little over these "discoveries," for we know where the Humanists discovered those classical texts—namely, in medieval manuscripts, where medieval scribes had been carefully preserving them for mankind over the centuries. … In view of all this disagreement over the duration of the Middle Ages, perhaps we should content ourselves with saying that our period extends from the close of the classical period to the beginning of the Renaissance. If classicists and Renaissance scholars don't know when their periods begin and end, then that is their problem.

Still, it is perhaps most useful not to think of medieval philosophy as defined by the chronological boundaries of its adjacent philosophical periods, but as beginning when thinkers first started to measure their philosophical speculations against the requirements of Christian doctrine and as ending when this was no longer the predominant practice.[3] This view allows late ancient and early medieval philosophy to overlap during the Patristic period; thus Proclus (411–85) belongs to the story of ancient philosophy, even though he is later than Saint Augustine (354–430). Again, this view accommodates the fact that late scholasticism survived and flourished even in the Renaissance. Thus Francisco Suárez (1548–1617), who can arguably be regarded as the last chapter in the history of medieval philosophy, was contemporary with Francis Bacon (1561–1626). Nevertheless by c. 1450, at the latest, radically new ways of doing philosophy were clearly emerging.

This perhaps generous interpretation of the chronological limits of medieval philosophy implies that it lasted at least from the Greek patristic author Justin Martyr (mid-second century) until well into the fifteenth century—more than half the entire history of philosophy generally. Clearly there is much to be discussed.

2. The Main Ingredients of Medieval Philosophy

Here is a recipe for producing medieval philosophy: Combine classical pagan philosophy, mainly Greek but also in its Roman versions, with the new Christian religion. Season with a variety of flavorings from the Jewish and Islamic intellectual heritages. Stir and simmer for 1300 years or more, until done.

This recipe produces a potent and volatile brew. For in fact many features of Christianity do not fit well into classical philosophical views. The notion of the Incarnation and the doctrine of the Trinity are obvious cases

in point. But even before those doctrines were fully formulated, there were difficulties, so that an educated Christian in the early centuries would be hard pressed to know how to accommodate religious views into the only philosophical tradition available. To take just one example, consider pagan philosophical theories of the soul. At first glance, it would appear that the Platonic[4] tradition would be most appealing to an early Christian. And in fact it was. In the first place, the Platonic tradition was very concerned with the moral development of the soul. Again, that tradition saw the highest goal of a human being as some kind of mystical gazing on or union with the Form of the Good or the One; it would be easy to interpret this as the "face to face" encounter with God in the next life that St. Paul describes in 1 Cor. 13:12. Most important of all, Platonism held that the soul could exist apart from the body after death. This would obviously be appealing to Christians, who believed in an afterlife.

On the other hand, there was another crucial aspect of Christianity that simply made no sense to a Platonist. This was the doctrine of the resurrection of the dead at the end of the world. Platonism allowed for reincarnation, so there was no special theoretical problem for a Platonist about the soul's reentering the body. But for a Christian this resurrection was something to *look forward to;* it was a *good* thing. This would be incomprehensible from a Platonic viewpoint, for which "the body is the prison of the soul," and for which the task of the philosopher is to "learn how to die" in order to be free from the distracting and corrupting influences of the body. No, for a Platonist it is best for the soul *not* to be in the body.[5]

A Christian would therefore have a hard time being a straightforward Platonist about the soul. But neither could a Christian be a straightforward Aristotelian. Aristotle's own views on the immortality of the soul are notoriously obscure, and he was often interpreted as denying it outright.

All the harder, therefore, to make sense of the view that the resurrection of the dead at the end of the world is something to be joyfully expected.[6]

This problem illustrates the kind of difficulties that emerge from the above "recipe" for medieval philosophy. Educated early Christians, striving to reconcile their religion in terms of the only philosophical traditions they knew, would plainly have a lot of work to do. Such tensions may be regarded as the "motors" that drove much of philosophy throughout the period. In response to them, new concepts, new theories, and new distinctions were developed. Of course, once developed, these tools remained and indeed still remain available to be used in contexts that have nothing to do with Christian doctrine. Readers of medieval philosophy who go on to study John Locke, for instance, will find it hard to imagine how his famous discussion of "personal identity" in the *Essay Concerning Human Understanding* could ever have been written if it were not for the medieval distinction between "person" and "nature," worked out in dealing with the doctrines of the Incarnation and the Trinity.

3. The Availability of Greek Texts

While the influence of classical pagan philosophy was crucial for the development of medieval philosophy, it is likewise crucial that until the twelfth and thirteenth centuries almost all the original Greek texts were lost to the Latin West, so that they exerted their influence only indirectly. They were "lost" not in the sense that the texts were simply unavailable but in the sense that very few people could read them, since they were written in the wrong language. As the Western Roman Empire gradually disintegrated, the knowledge of Greek all but disappeared. Boethius (c. 480–545/526) was still fluent in Greek, but he recognized the need for translations even in his own day; after him Greek was effectively a dead language in the West. There were still some pockets of Greek literacy, especially around such figures as Isidore of Seville and the Venerable

Bede, preserving and transmitting ideas of ancient learning, but making little impact on medieval philosophical thought.

In the case of Plato, the Middle Ages for all practical purposes had only the first part of the *Timaeus* (to 53c), hardly a typical Platonic dialogue, in a translation and commentary by a certain Calcidius (or Chalcidius).[7] The *Timaeus* contains Plato's cosmology, his account of the origin of the cosmos.

There were also translations of the *Meno* and the *Phaedo* made in the twelfth century by a certain Henry Aristippus of Catania,[8] but almost no one appears to have read them. They seem to have had only a modest circulation and absolutely no influence at all to speak of.[9]

There had been a few other Latin translations made even much earlier, but these vanished from circulation before the Middle Ages got very far along. Cicero himself had translated the *Protagoras* and a small part of the *Timaeus*, and in the second century Apuleius translated the *Phaedo*, but these translations disappeared after the sixth century and had very little effect on anyone (Klibansky [1982], pp. 21–22). As Saint Jerome remarks in the late-fourth or early-fifth century, in his *Commentary on the Epistle to the Galatians*, "How many people know Plato's books, or his name? Idle old men on the corners hardly recall him" (Migne [1844–64], vol. 26, col. 401B).

This state of affairs lasted until the Renaissance, when Marsilio Ficino (1433–99) translated and commented on the complete works of Plato. Thus, except for roughly the first half of the *Timaeus*, the Middle Ages did not know the actual texts of Plato.

As for Plotinus, matters were even worse. His *Enneads* (the collection of his writings) were almost completely unavailable. Marius Victorinus is said to have translated some of the *Enneads* into Latin in the fourth

century, but his translation, if in fact it really existed, seems to have been lost soon afterwards.[10]

For Aristotle, the Middle Ages were in somewhat better shape. Marius Victorinus translated the *Categories* and *On Interpretation*. A little over a century later, the logical works in general, except perhaps for the *Posterior Analytics*, were translated by Boethius, c. 510–12, but only his translations of the *Categories* and *On Interpretation* ever got into general circulation before the twelfth century. The rest of Aristotle was eventually translated into Latin, but only much later, from about the middle of the twelfth century. First there came the rest of the logical works, and then the *Physics*, the *Metaphysics*, and so on. Essentially all the works had been translated by the middle of the thirteenth century (Dod [1982]). This "recovery" of Aristotle in the twelfth and thirteenth centuries was a momentous event in the history of medieval philosophy.

Still, while it is important to emphasize this absence of primary texts of Greek philosophy in the Latin Middle Ages, it is also important to recognize that the medievals knew a good deal about Greek philosophy anyway. They got their information from (1) some of the Latin patristic authors, like Tertullian, Ambrose, and Boethius, who wrote before the knowledge of Greek effectively disappeared in the West, and who often discuss classical Greek doctrines in some detail; and (2) certain Latin pagan authors such as Cicero and Seneca, who give us (and gave the medievals) a great deal of information about Greek philosophy.

During the first part of the Middle Ages, Platonic and neo-Platonic influences dominated philosophical thinking. "Plato himself does not appear at all, but Platonism is everywhere," as Gilson has said. (Gilson [1955], p. 144.[11]) This situation prevailed until the recovery of Aristotle in the twelfth and thirteenth centuries. Hence, even though it is sometimes still done, it is quite wrong to think of medieval philosophy as mainly just

a matter of warmed-over commentaries on Aristotle. For most of the Middle Ages by far, Aristotle was of decidedly secondary importance. This of course is not to deny that when Aristotle did come to dominate, he was *very* dominant indeed and his influence was immense.

4. From the Patristic Period to the Mid-Twelfth Century

"Patrology" or "patristics" is the study of the so called "Fathers (*patres*) of the Church." In this sense, 'fathers' does not mean priests, although of course many patristic authors *were* priests. Neither does it does mean "fathers" in the sense of "founding fathers," although many patristic authors were likewise foundational for everything that came afterward. Rather 'fathers' in this sense means "teachers." See, for example, St. Paul: "For though you might have ten thousand guardians in Christ, you do not have many fathers. Indeed, in Christ Jesus I became your father through the gospel" (1 Cor. 4:15 [NRSV]). In early Christian usage, the term 'father' was applied primarily to the bishop, who had preeminent teaching authority within the Church. But gradually the word was extended until, much later, it came to include all early Christian writers who were taken to represent the authentic tradition of the Church (Quasten [1950–86], I, p. 9). The patristic period is generally taken to extend from the immediately post-Apostolic authors to either Gregory the Great (d. 604) or Isidore of Seville (d. 636) in the Latin West, and to John of Damascus (d. 749) in the Greek East (Quasten [1950–86], I, 1).

4.1 Augustine

By no means all patristic authors are of philosophical significance, but many of them definitely are. By far the most important is Saint Augustine (354–430) (see the entry on Saint Augustine). Augustine is certainly the

most important and influential philosopher of the Middle Ages, and one of the most influential philosophers of any time:[12]

> His authority has been felt much more broadly, and for a much longer time, than Aristotle's, whose role in the Middle Ages was comparatively minor until rather late. As for Plato, for a long time much of his influence was felt mainly through the writings of Augustine. For more than a millennium after his death, Augustine was an authority who simply had to be accommodated. He shaped medieval thought as no one else did. Moreover, his influence did not end with the Middle Ages. Throughout the Reformation, appeals to Augustine's authority were commonplace on all sides. His theory of illumination lives on in Malebranche and in Descartes's "light of nature." His approach to the problem of evil and to human free will is still widely held today. His force was and is still felt not just in philosophy but also in theology, popular religion, and political thought, for example in the theory of the just war. (Spade [1994], pp. 57–58)

Yet despite his philosophical preeminence, Augustine was not, and did not think of himself as, a philosopher either by training or by profession. By training he was a rhetorician, by profession first a rhetorician and teacher of rhetoric, then later Bishop of Hippo (modern Annaba, or French Bône, in what is now northeast Algeria), where his concerns were pastoral and theological. As a result, few of his writings contain what we would think of as purely philosophical discussions.[13] What we find instead in Augustine is a man who is a "philosopher" in the original, etymological sense, a "lover a wisdom," one who is *searching* for it rather than one who writes as if he has found it and is now presenting it to us in systematic, argumentative form.

4.2 Boethius

After Augustine, the first thinker of philosophical note was Boethius (c. 480–524/525) (see the entry on Anicius Manlius Severinus Boethius). Boethius is no doubt best known today for *The Consolation of Philosophy*, a dialogue in five books between Boethius and "Lady Philosophy," an allegorical figure who appears to him in a vision while he is languishing in jail under sentence of death for treason. Boethius had occupied a high station in society and government. He was born into a family with an excellent old Roman pedigree, and rose to a position of immense power and influence in the Ostrogothic kingdom under Theodoric. Although for a while he was conspicuously successful, he nevertheless eventually fell into disfavor, was charged with treasonable conspiracy having to do with the Emperor Justin in Constantinople (Boethius claims he was innocent), was arrested and finally executed.[14] In the *Consolation*, Boethius and Lady Philosophy discuss the problem of evil and the fickleness of fortune—a particularly pressing issue for Boethius, given the circumstances under which the work was written.

But although the *Consolation* is justly famous, both in our own day and in the Middle Ages, Boethius's long-term importance probably rests more on his translations and commentary activity. For Boethius was well educated, and was one of the increasingly rare people in the West who knew Greek well, not just the language but the intellectual culture. He came up with the lofty goal to translate Plato and Aristotle into Latin, write commentaries on the whole of that material, and then write another work to show that Plato and Aristotle essentially said the same thing:

> If the more powerful favor of divinity grants it to me, this is [my] firm purpose: Although those people were very great talents whose labor and study translated into the Latin tongue much of what we are now treating, nevertheless they did not bring it into any kind of order or shape or in its arrangement to the level of the [scholarly] disciplines. [Hence I propose] that I turn all of Aristotle's work—

[or] whatever [of it] comes into my hands—into the Latin style and write commentaries in the Latin language on all of it, so that if anything of the subtlety of the logical art was written down by Aristotle, of the weightiness of moral knowledge, of the cleverness of the truth of physical matters, I will translate it and even illuminate it with a kind of "light" of commentary. [Then,] translating all of Plato's dialogues or even commenting [on them], I will bring them into Latin form. Once all this is done, I will not fail to bring the views of Aristotle and Plato together into a kind of harmony and show that they do not, as most people [think], disagree about everything but rather agree on most things, especially in philosophy. (Boethius [1880], pp. 79. 9–80.6 [my translation])

No doubt this plan would have proved unmanageable even if Boethius had not been executed in his mid-forties. In particular, while the *Consolation* certainly shows a knowledge of the *Timaeus*, Boethius does not appear to have actually translated *any* Plato at all, despite his intentions. He did, however, translate Aristotle's *Categories* and *On Interpretation*, together with Porphyry of Tyre's *Isagoge*, a kind of "introduction" to Aristotle's *Categories*.[15] He also appears to have translated the other works in Aristotle's *Organon* (except perhaps for the *Posterior Analytics*, about which there is some doubt), but the fate of those translations is obscure; they did not circulate widely until much later (Dod [1982], pp. 53–54).

In addition to his translations, Boethius wrote a number of logical treatises of his own. These are, first of all, a commentary on Aristotle's *Topics*, which is no longer extant. Whether or not he translated the *Posterior Analytics*, there may have been a commentary on it, but if so it has not survived and did not have any influence (Ebbesen [1973]). The same goes for a possible (incomplete) commentary on the *Prior Analytics* (Obertello [1974], I, pp. 230–32). More important were a series of commentaries

(one on the *Categories*, two each on *On Interpretation* and on Porphyry's *Isagoge*, and one on Cicero's *Topics*) (see the entry on medieval theories of categories), together with several other works on categorical and hypothetical syllogisms, logical "division," and on the differences between Aristotle's and Cicero's *Topics* (Chadwick [1981], Gibson [1981], Obertello [1974]). Together all these logical writings, both the translations and the others, constitute what later came to be called the "Old Logic" (= *logica vetus*). Some of the works were more influential than others. But basically, everything the Middle Ages knew about logic up to the middle of the twelfth century was contained in these books. As a result, Boethius is one of the main sources for the transmission of ancient Greek philosophy to the Latin West during the first half of the Middle Ages.

Boethius is also important for having introduced the famous "problem of universals" in the form in which it was mainly discussed throughout the Middle Ages (see the entry on the medieval problem of universals).

He also proved to be influential in the twelfth century and afterwards for the metaphysical views contained in a series of short studies known collectively as the *Theological Tractates*.

4.3 The Carolingian Period

After Boethius, as the classical Greco-Roman world grew ever more distant, philosophy—and to some extent culture generally—entered a period of relative stagnation, a period that lasted until after the year 1000. There was one short-lived bright spot, however, the late-eighth and early-ninth century court of Charlemagne (768–814) and his successors, the so called "Carolingian" period. The major philosophical figure in this period was John Scottus Eriugena[16] (c. 800–c. 877), an Irish monk who was at the court of Charles the Bald around 850 (see the entry on John Scottus Eriugena). Curiously, the knowledge of Greek was still not quite dead in

Ireland even at this late date, and Eriugena brought a knowledge of the language with him. At the Carolingian court, Eriugena translated several Greek works into Latin, including the very important writings of Pseudo-Dionysius the Areopagite (more on him below), a work by Maximus Confessor (also known as Maximus of Constantinople, c. 580–662), and Gregory of Nyssa's (died c. 385) *On the Making of Man* (= *De hominis opificio*). Eriugena also wrote several other works of his own.

Among his translations, the writings of Pseudo-Dionysius are surely the most important and influential (see the entry on Pseudo-Dionysius the Areopagite).[17] The true identity of the man we call "Pseudo-Dionysius" is unknown, but he lived probably in the late-fifth century, somewhere in the Greek-speaking near East, and was very much influenced by the late neo-Platonist Proclus. Whoever he was, he *claimed* to be a certain Dionysius who is reported to have been among the philosophers on the Areopagus in Athens when St. Paul went there to preach (Acts 17:19–34). Most of the audience on that occasion laughed at Paul and his novel doctrines.

> But some of them joined him and became believers, including Dionysius the Areopagite and a woman named Damaris, and others with them. (Acts 17:34)

Damaris and the "others" have disappeared without a trace, but our unknown later author *pretends* to be the Dionysius mentioned in this passage.

The Pseudo-Dionysian writings consist of four treatises and a series of ten letters. The most philosophically important of them are the two treatises *On the Divine Names* and *On Mystical Theology*. Through them the Latin West was introduced to what is sometimes called "darkness mysticism," the tradition that interprets mystical experience not in terms of an

"intellectual vision" (compare Plato's Allegory of the Cave, where the Form of the Good is described as the dazzling sun), but in terms of the will rather than the intellect, darkness rather than light. (Compare later mystical expressions such as "dark night of the soul," "cloud of unknowing.")

It is also mainly through these two treatises that medieval philosophy got the still familiar view that there are three ways of talking about God, by trying to say what he is like (the *via affirmativa*), by saying instead what he is *not* (the *via negativa*), and by a kind of "combined" way that speaks of God with affirmative predicates, but with some kind of mark of superexcellence (the *via eminentiae*, "God is *more than* good, *more than* wise.").

Among Eriugena's own writings, the two most important ones were surely *On the Division of Nature* (= *De divisione naturae* or, under a Greek title, *Periphyseon*) and *On Predestination* (= *De praedestinatione*), both very strongly influenced by the neo-Platonic texts Eriugena was translating. Both works were condemned, *On Predestination* soon after it was written. *On the Division of Nature* is a large, systematic work in four books, presenting a vision of reality in strongly neo-Platonic terms. The unfamiliarity of this kind of thinking in Western Christendom, which was strongly influenced by Augustine, no doubt contributed to his later reputation of being a heretic.

4.4 Anselm of Canterbury

After its brief "renaissance" during the Carolingian period, education and culture declined once again for roughly another 200 years. Then, shortly after the turn of the millennium, things began to revive. The Germanic "barbarian" tribes that had so disturbed the late Roman empire had long since settled down, and the later Viking raiders had by this time become

respectable "Normans." Trade began to revive, travel became relatively safe again, at least compared to what it had been, new cities began to emerge, and along with them new social arrangements began to develop. Education was part of this general revival, and with it philosophy. The major medieval philosophers before the year 1000 are probably fewer than five in number (depending on how generously one wants to take the word 'major'). But after 1000 their numbers grow exponentially. It is no longer possible to treat them individually in chronological order; indeed, it is difficult to keep track of them all. As time goes on, the complications and the numbers only increase.

Simultaneously, philosophy becomes increasingly technical and "academic." Anselm of Canterbury (1033–1109) represents a major transitional figure (see the entry on Saint Anselm). His writings are not yet laden with the technicalities and jargon that make so much later medieval philosophy formidable and inaccessible to the non-specialist. And yet his writings are philosophically "argumentative" in a way much earlier medieval philosophy is not and that looks much more familiar to present-day readers.

Anselm is no doubt best known as the originator of the famous "ontological argument" for the existence of God.[18] But he wrote much else besides, on many philosophical and theological topics. His writings abound in subtle and sophisticated reasoning; indeed, they illustrate the increasing role of "dialectic" in philosophy and theology. In Anselm's hands, theology begins to develop into an argumentative discipline, less exclusively a matter of "scripture studies" and spirituality and increasingly a matter of systematic exploration and presentation of doctrine. This development grows even more pronounced after Anselm.

4.5 Peter Abelard

By the early twelfth century, the revival of education that had begun shortly after the millennium was in full swing. During the first half of the century, the most important philosopher by far was undoubtedly Peter Abelard (1079–1142) (see the entry on Peter Abelard). He was also one of the most colorful figures in the entire history of philosophy. His affair with Héloise and his consequent castration are the stuff of legend, and his controversy with the much more traditional Bernard of Clairvaux (1090–1153) has only enhanced his reputation among those who have viewed him (with considerable oversimplification) as a champion of reason over authority. His autobiographical *Story of My Adversities* (= *Historia calamitatum*) is a "good read" even today, and is one of the most intensely personal documents of the Middle Ages.[19]

Abelard represents the full flower of "early medieval philosophy," just before the new translations of Aristotle and others transform everything. It is important to realize that, except for the works of Pseudo-Dionysius, which do not appear to have had an important role in Abelard's thinking, he had access to no more of the original sources of philosophy in the ancient world than anyone else in Europe had had since the time of Boethius. Yet his philosophy is strikingly original. His views on logic and what we would call philosophy of language are sophisticated and novel; indeed, he is a serious contender for the title of the greatest logician of the entire medieval period, early or late. He is one of the first nominalists, and certainly the first important one. His writings on ethics put a new and very strong emphasis on the role of the agent's *intention* rather than exterior actions. He also wrote on theological topics such as Trinity.

Abelard's writings further amplify the tendency, already seen in Anselm, to increase the use of reasoning and argumentation in theology. But whereas Anselm had managed to deflect criticisms of this new approach in theology, Abelard's disputatious personality alarmed those who were more comfortable with the older style. He was subject to ecclesiastical censure

during his lifetime, a fact that no doubt contributes to the relatively few explicit citations of him in the later Middle Ages. Nevertheless, it is undeniable that his influence was widespread.

4.6 General Characteristics of This Early Period

Throughout this early medieval period, we find many writers, usually of a broadly "Platonic" persuasion, who deal with philosophical topics in an unsystematic but far from shallow way that does not clearly distinguish philosophy from theology, or for that matter from "wisdom literature" generally. Frequently their views are presented by arguments that amount to an appeal to a "vision" of how things are ("Look, don't you see?").[20] This is simply a general although not universal observation about these authors, and should not be regarded as a philosophical limitation or defect. After all, some of the world's most important philosophy has been presented in such a "visionary" way. Consider the role of "intuition" in twentieth-century phenomenology, for example, not to mention Parmenides's poem (where the philosophy is presented by a goddess) and much of Plato's philosophy, including the Allegory of the Cave.

There are many exceptions to this generalization. Boethius's logical commentaries, for example, are purely philosophical and frequently genuinely argumentative, even if they are often obscure and inaccessible to modern readers. Eriugena's *On the Division of Nature*, while definitely "visionary," is nevertheless quite systematic in its structure. And by the time of Anselm, the role of logical argumentation is beginning to grow. Certainly for Abelard the above generalization fails entirely.

Nevertheless, a big change is about to occur. Prior to Abelard, philosophy in the Middle Ages had not been an exclusively academic affair. It had been addressed for the most part to any well educated reader interested in the topics being discussed. Boethius's *Consolation*, for instance, or almost

any of Augustine's or Anselm's writings, could profitably be read by any literate person. Soon, however, this all changes. Philosophy becomes an increasingly specialized discipline, pursued by and for those whose livelihood is found only in educational institutions. Philosophy and theology become more clearly distinguished from one another; both become more systematic, rigorous and precise. These virtues are accompanied by an increasingly technical jargon, which makes so much late-medieval philosophy intimidating and formidable to non-specialist readers. By the same token, this increasing technicality diminishes the overall sense of moral urgency one finds for example in Augustine's *Confessions* or Boethius's *Consolation*.

As with the previous generalization, this one should not be regarded as a philosophical fault of the later authors; it is simply a different way of doing philosophy. As David Hume knew, there are two styles of philosophy, each with its own advantages (*An Enquiry Concerning Human Understanding*, § 1). What we see in passing from the earlier to the later Middle Ages is a transition from one to the other.

5. The Twelfth Century and the Rise of Universities

5.1 New Translations[21]

As part of the cultural revival described above, and from the late-eleventh century on, there was a new and increasing interest in having translations of previously unavailable texts, not all of them philosophical by any means. No doubt this new interest was prompted in part by Western Europe's exposure to the Greek and Islamic world during the First Crusade (beginning in 1095). But, for whatever reason, new translations soon began to appear from:

- Sicily, which was at this time a melting-pot of Latins, Greeks, Jews,

and Muslims. Euclid and Ptolemy were translated there, as well as other mathematical and medical works.

- Constantinople. A few Western scholars journeyed to Constantinople, notably one James of Venice in roughly the late 1120s, an important translator of Aristotle's logical and other writings. Nevertheless, political tensions between the West and Constantinople at this time guaranteed that such contact was not widespread (see the entry on Byzantine philosophy).

- Spain. An extremely important school of translators emerged at Toledo, under the direction of Archbishop Raymond (d. 1151, although the school survived him). They included, among others:

  - John of Spain (Johannes Hispanus) who translated, among other things, the immensely important Muslim philosopher Avicenna's (Ibn Sina, 980–1037) *Logic* from Arabic into Latin.

  - Dominic Gundissalinus (or Gundisalvi, an old form of "Gonzales," fl. late-twelfth century). Gundissalinus translated Avicenna's *Metaphysics*, part of his *Physics*, and some of his other works, as well as writings by the Islamic philosophers Al-Farabi (c. 870–950) and Al-Ghazali (1058–1111). Together with John of Spain, Gundissalinus translated Solomon Ibn Gabirol's (c. 1022–c. 1058/c.1070) *Fountain of Life* (= *Fons vitae*). Ibn Gabirol (in Latin, Avicebron, Avencebrol, etc.) was an Iberian Jewish author whose *Fountain of Life* was written in Arabic. It presents a systematic neo-Platonic view of the cosmos. In addition to these translations, Gundissalinus was also the author of some original philosophical works of his own.

  - Gerard of Cremona (d. 1187). Gerard began work at Toledo in 1134. He translated Aristotle's *Posterior Analytics*, together with Themistius's commentary on it, Aristotle's *Physics, On the Heavens, On Generation and Corruption*, and parts of his *Meteorology*, the Muslim Al-Kindi's (d. 873) important *On the*

> *Intellect* and other works of his. Gerard also translated the very
> important *Book of Causes* (= *Liber de causis*), falsely attributed
> to Aristotle although the work is in fact based on certain theses
> extracted from Proclus's *Elements of Theology*.

The Spanish translators worked from Arabic texts. In the case of Aristotle, they used Arabic translations of Aristotle's Greek, sometimes with an earlier Syriac link in between. After such a circuitous route, it is no less than amazing that the Latin Europeans were able to understand anything at all of these newly available Aristotelian works. Eventually the extensive and thorough commentaries by the Moorish Ibn Rushd (in Latin, Averroes, 1126–98) were translated from Arabic as well. These commentaries were extremely important in shaping the late medieval understanding of Aristotle, although some of the views contained in them became highly controversial.

By the end of the twelfth century, almost all of Aristotle's works available today had been translated into Latin and, together with the commentaries and other newly translated texts, gradually began to circulate. By the mid-thirteenth century, they were widely known. The first things to spread were the remaining logical writings of Aristotle's *Organon*, those not already widely known from Boethius's translations some six hundred years previously. These new logical writings, as distinct from the "Old Logic" (= *Logica vetus*) stemming from Boethius, became known collectively as the "New Logic" (= *Logica nova*). After them, the *Physics, Metaphysics* and other Aristotelian writings gradually became known.

This relatively sudden injection of so much new and unfamiliar material into Western Europe was a stunning shock, nothing less than revolutionary. It was no longer possible for philosophers and theologians to regard their task as simply one of deepening and elaborating traditional views that had come mainly from the Church Fathers and other familiar

and approved authorities. It was now a matter of dealing with an entirely unfamiliar framework, with new ideas, accompanied by powerful arguments for them, some of which ideas were plainly unacceptable to a Christian—for example, Aristotle's rejection of anything like divine providence, and his views on the eternity of the world (see the entry on William of Auvergne).

5.2 New Forms of Education

As part of the revival that began after the turn of the millennium, new forms of education began to emerge in Western Europe. In general, we may distinguish four main types of educational practices in the Middle Ages:[22]

- **Monastic schools.** These were schools that had been regularly associated with monasteries ever since the sixth century. Much of Anselm's most important work, for instance, including the *Proslogion* containing his "ontological argument," was penned at the monastic school of Bec in Normandy. Abelard in his *Story of My Adversities* describes how, at least according to Abelard's telling, his teacher William of Champeaux (c. 1070–1121) was driven out of Paris by Abelard's superior dialectical skills and retired to the abbey of Saint Victor, where he "founded" (or at least reorganized) what came to be known as the School of Saint Victor. This was another one of these monastic schools. The masters of this school became quite well known in their own right in the later-twelfth century. They are collectively known as the "Victorines." The most important of them are:

  - Hugh of St. Victor (c. 1096–1141), the author of a *Didascalicon* on the various liberal arts. Hugh was also a theologian and theorist of mysticism.

- Richard of St. Victor (c. 1123–73), who succeeded Hugh as master of the school. Richard, like Hugh, was a theorist of mysticism. He also wrote an important treatise on the doctrine of the Trinity, the first serious alternative to Augustine's approach in the latter's own *On the Trinity*. Unlike Hugh, Richard was much more favorably disposed toward the new use of dialectic or logic in theology. He is said to have written a treatise of his own on logic but it does not appear to have survived.

- **Individual "masters."** Beginning in the mid-eleventh century, individual scholars would occasionally set up a "school" of their own and gather students around them. Such schools were sometimes itinerant, and depended entirely on the appeal of the teaching "master." Perhaps the closest analogue to this arrangement would be the modern "martial arts" schools one often finds in present-day cities. The practice declined after c. 1150. Abelard conducted such a "school" at Melun in the very early eleventh century, and seems earlier to have attended a similar "school" conducted by a certain Roscelin (c. 1045–c. 1120), a controversial nominalist whose writings have mostly not survived, but who had in effect been accused by Anselm of out-and-out tritheism on the doctrine of the Trinity.

- **Cathedral schools.** These were schools associated with the official church of a bishop, and played a role similar to that of the monastic schools for monasteries: they trained young clerics and occasionally others as well. Before William of Champeaux left Paris as the result of Abelard's criticisms of his views, he had been teaching at the cathedral school of Paris (see the entry on William of Champeaux). The so-called "School of Chartres" may likewise have been such a cathedral school.[23] The scholars there were especially interested in that portion of Plato's *Timaeus* that was circulating in Calcidius's translation (see above), and in the metaphysical implications of

Boethius's *Theological Tractates*. Important figures associated with the School of Chartres include Bernard of Chartres (died c. 1130), Thierry (= Theodoric) of Chartres (died c. 1150), and Gilbert of Poitiers (= Gilbert de la Porrée, Gilbertus Porreta, c. 1085–1154). John of Salisbury's (c. 1115–80) *Metalogicon* is an invaluable source of information about all these and many other thinkers from the first half of the twelfth century (John of Salisbury [1955]; see John of Salisbury).[24] Cathedral schools flourished c. 1050–c. 1150.

- **Universities.** Parliament and the "university" are arguably the two great medieval institutions that have survived more or less intact to the present day. (The Church may be counted as a conspicuous third, depending on one's views about the Reformation and Counter-Reformation.) Frequently, universities grew out of cathedral schools. Thus, the cathedral school at Paris developed by the early-thirteenth century into the University of Paris. An important cathedral school drew students from all over Europe. Such a school became known as a *studium generale*. Some of these *studia generalia* survived and became known as "universities." At first, the term 'universitas' referred simply to the "entirety" or "universality" of scholars, both faculty and students, associated with the school. As the term gradually came to be used, a "university" was one of these major, international schools that was distinguished from others by its possessing an official charter (granted by a royal or ecclesiastical authority), a set of statutes, and an established form of governing itself.

The University of Paris was the premier university in Europe in the thirteenth century. Its statutes were officially approved by the papal legate Robert de Courçon in 1215. The official founding of the University is usually put at this date, although it is clear that the statutes existed earlier. Oxford and Cambridge also date from the early-thirteenth century,

although their period of greatest vigor in the Middle Ages came in the late-thirteenth and early-fourteenth century. Toulouse was founded in 1229 by papal charter. Salamanca was founded by royal charter in 1200. There were also universities in Italy; indeed, Bologna was the first university in all of Europe, and had the peculiarity of being a *student*-run university.

Universities were divided into "faculties." The four most common ones were the faculties of arts, law, medicine, and theology. Most universities had arts faculties, in addition to one or more of the others. The arts faculty was for the basic training of students, before they proceeded to one of the "higher" faculties. In effect, the arts faculty was the equivalent of the modern undergraduate program. As for the "higher" faculties, Bologna was primarily a university for the study of law. Others were best known for medicine. Paris had all four faculties, but the faculty of theology was considered the highest of the four.

In the medieval university, philosophy was cultivated first and foremost in the arts faculty. When the newly translated works of Aristotle first appeared at the University of Paris, for instance, it was in the faculty of arts. The works were clearly not law or medicine. (Some of them might be stretched a bit to count as medicine, but these were not the ones that were influential first.) Neither were they theology in the traditional sense of "Sacred Doctrine," although some of Aristotle's writings had important consequences for theology. Some of these consequences were thought to be dangerous for Christian doctrine, and they were. In 1210, a provincial synod at Paris ruled that Aristotle's "natural theology" could not be "read" in the faculty of arts at Paris. To "read" in this context means to "lecture on." It did not mean that students and masters couldn't study and discuss these works in private. In 1215, when Robert de Courçon approved the statutes of the University of Paris, one of them forbade the arts masters from lecturing on Aristotelian metaphysics and natural science. In 1231, Pope Gregory IX ordered that the works prohibited in 1210 not be used

until they could be examined by a theological commission to remove any errors. In 1245, Innocent IV extended the prohibitions of 1210 and 1215 to the University of Toulouse. Despite these bans, study and discussion of Aristotle could not be stopped. By the 1250s, people were openly lecturing on everything they had of Aristotle's.

Why were these prohibitions issued? In part it was out of a genuine concern for the purity of the faith. Aristotelianism was thought, and rightly so, to be theologically suspect. Besides, European academics were just getting acquainted with most of Aristotle, and at this early stage of their acquaintance they weren't altogether sure just what he meant and what the implications were. A "go slow" approach was not an altogether unreasonable course of action to adopt. On the other hand, it cannot be denied that some of the basis for the prohibitions was simply a resistance to new ideas.

6. The Thirteenth Century and Later

By their very nature, universities brought together masters and students from all over Europe and put them in close proximity. Not surprisingly, the result was a "boom" in academic study, including philosophy. Already in the twelfth century, and certainly by the early-thirteenth, it is futile even to attempt anything like a sequential narrative of the history of medieval philosophy. Instead, the remainder of this article will mention only a few of the major figures and describe some of the main topics that were discussed throughout the medieval period. For a more complete picture, readers should consult any of the general histories in the Bibliography below, and for details on individual authors and topics the Related Entries in this Encyclopedia, listed below.

Histories of medieval philosophy often treat Thomas Aquinas (1224/25–74), John Duns Scotus (c. 1265–1308), and William of Ockham (c. 1287–

1347) as the "big three" figures in the later medieval period; a few add Bonaventure (1221–74) as a fourth. Although there is certainly ample justification for giving special emphasis to these authors, it would be misleading if one thought one could get even a fair overall picture from them alone. Nevertheless, the list is instructive and illustrates several things.

First of all, not one of these three or four authors was French. Aquinas and Bonaventure were Italian, Scotus—as his name implies—was a Scot, and Ockham was English. All but Ockham spent at least part of their careers at the University of Paris. This illustrates both the preeminence of the University of Paris in the thirteenth century and the increasing internationalization of education in the later Middle Ages in general. But it also illustrates another odd fact: the relative absence of Frenchmen as major players on the philosophical scene during this period, even at the premier university in France. There are certainly notable exceptions to this perhaps contentious observation (see for example the entries on Peter Auriol, John Buridan, Godfrey of Fontaines, Nicholas of Autrecourt, Peter John Olivi, Philip the Chancellor, and William of Auvergne), but with the arguable exception of Buridan, surely none of them is of the stature of the four mentioned above.

The fact that Buridan has not been generally acknowledged in the same rank as the four "greats," even though he is certainly a formidable contender, points to an important feature of the twentieth-century historiography of later medieval philosophy. Buridan was what is known as a "secular master." That is, although he was a priest, he did not belong to any of the religious "orders."[25] Beginning in the early-thirteenth century, several new orders were founded, notably the Franciscans (1209) and the Dominicans (1216), both of which became very prominent in late medieval universities. Aquinas was a Dominican, while Bonaventure, Scotus, and Ockham were Franciscans.

Religious orders tend to keep good records, including the writings of their members, so that historians of medieval philosophy typically have more material to work with for authors in the various orders than they do for "secular" figures like Buridan. Besides, other things being equal, orders understandably prefer to "champion their own" in academic as in other matters, and when the academic champion comes relatively early in the history of his order, he can come to be regarded as representing the order's authentic "position," thereby influencing the views of later members of the order.[26] In this way, Aquinas soon became the semi-"official" philosopher and theologian of the Dominicans, a status that was enhanced in 1879 in Pope Leo XIII's encyclical *Aeterni Patris*, which called Aquinas "the chief and master of all the scholastic doctors," and urged that preference be given to Thomistic doctrine in Catholic schools (see the entry on Saint Thomas Aquinas). As a result, Aquinas enjoyed a far greater authority in the late-nineteenth and the first half of the twentieth century than perhaps he ever did in the Middle Ages. To some extent, Bonaventure likewise came to be regarded as representing typically Franciscan views (see the entry on Saint Bonaventure), and later on Scotus was highly respected and often favored among the Franciscans (see the entry on John Duns Scotus). Ockham is a special case. He was a controversial figure, mainly because of political disputes with the Pope that embroiled his later life (see the entry on William of Ockham). Nevertheless, as one of their own, the Franciscans have always been interested in him and in his writings.

The upshot of all this is that major late medieval philosophers, like Buridan, who did not belong to a religious order have often suffered from neglect in standard histories of medieval philosophy, at least until fairly recently. Another neglected secular master was Henry of Ghent, a very important late-thirteenth century figure who has turned out to be crucial for understanding much of Duns Scotus, but whose views have only in the

last few decades begun to be seriously studied (see the entry on Henry of Ghent).

For that matter, even many important and influential late medieval philosophers who did belong to religious orders are still virtually unknown or at least woefully understudied today, despite the labors of generations of scholars. Their works have never been printed and exist only in handwritten manuscripts, written in a devilishly obscure system of abbreviation it takes special training to decode. It is probably safe to say that for no other period in the history of European philosophy does so much basic groundwork remain to be done.

7. Some Main Topics in Medieval Philosophy

Medieval philosophy included all the main areas we think of as part of philosophy today. Nevertheless, certain topics stand out as worthy of special mention. To begin with, it is only a slight exaggeration to say that medieval philosophy invented the philosophy of religion. To be sure, ancient pagan philosophers sometimes talked about the nature of the gods. But a whole host of traditional problems in the philosophy of religion first took on in the Middle Ages the forms in which we still often discuss them today:

- The problem of the compatibility of the divine attributes.
- The problem of evil. Ancient philosophy had speculated on evil, but the particularly pressing form the problem takes on in Christianity, where an omniscient, omnipotent, and benevolent God freely created absolutely everything besides himself, first emerged in the Middle Ages.
- The problem of the compatibility of divine foreknowledge with human free will. Many medieval authors appealed to human free will in their response to the problem of evil, so that it was especially

important to find some way to reconcile our free will with divine foreknowledge (see the entry on medieval theories of future contingents).

As for logic, the great historian of logic I. M. Bocheński ([1961], pp. 10–18) remarked that the later Middle Ages was—along with the ancient period from roughly 350–200 BCE and the recent period from Boole and Peano on—one of the three great, original periods in the history of logic. Although we have learned much about the history of logic since Bocheński wrote, and although we can find individual notable figures in logic who fall outside any of his three great periods, his observation is still by and large correct. From the time of Abelard through at least the middle of the fourteenth century, if not later, the peculiarly medieval contributions to logic were developed and cultivated to a very high degree. It was no longer a matter of interpreting Aristotle, or commenting on the works of the "Old Logic" or the "New Logic"; wholly new genres of logical writing sprang up, and entirely new logical and semantic notions were developed. For logical developments in the Middle Ages, see the articles insolubles, literary forms of medieval philosophy, medieval theories of categories, medieval semiotics, medieval theories of analogy, medieval theories of demonstration, medieval theories of modality, medieval theories of *Obligationes*, medieval theories: properties of terms, medieval theories of singular terms, medieval theories of the syllogism, and sophismata. For information on some contributors to medieval logic, see the articles Albert of Saxony, Anicius Manlius Severinus Boethius, John Buridan, John Wyclif, Johannes Sharpe, Paul of Venice, Peter Abelard, Peter of Spain, Richard Kilvington, Richard the Sophister, Roger Bacon, Thomas of Erfurt, Walter Burley, William Heytesbury, and William of Ockham.

In metaphysics, the Middle Ages has a well deserved reputation for philosophical excellence. The problem of universals, for example, was one of the topics that were discussed at this time with a level of precision and

rigor it would be hard to find matched before or since. But it was by no means the only such question. For some of the main topics in metaphysics on which medieval philosophers sharpened their wits, see the articles *binarium famosissimum*, existence, medieval mereology, the medieval problem of universals, medieval theories of causality, medieval theories of haecceity, and medieval theories of relations. For some important contributors to medieval metaphysics, see the articles John Buridan, John Duns Scotus, John Wyclif, Saint Augustine, Saint Thomas Aquinas, and William of Ockham.

In natural philosophy and philosophy of science, medieval philosophy was of course very strongly—but not exclusively—influenced by Aristotle. See, for example, the articles medieval theories of causality and Saint Thomas Aquinas. Particularly from the fourteenth century on, the increasing use of mathematical reasoning in natural philosophy would eventually pave the way for the rise of early modern science later on. Important figures in this development include William Heytesbury and William of Ockham. Other important contributors to medieval natural philosophy include Albert of Saxony, Dietrich of Freiberg, John Buridan, Nicholas of Autrecourt, Nicole Oresme, Robert Grosseteste, and William Crathorn.

Medieval epistemology was not, with some noteworthy exceptions, particularly worried over the problem of skepticism, over *whether* we have genuine knowledge (see the entry on medieval skepticism). The tendency was to take it for granted that we do, and instead to ask about *how* this comes about: what are the mechanisms of cognition, concept formation, etc. Medieval epistemology, therefore, typically shades into what we would nowadays call philosophical psychology or philosophy of mind; after the recovery of Aristotle's *On the Soul*, it was regarded as a branch of the philosophy of nature. For some of the important topics discussed in the area of medieval epistemology, see the entries divine illumination,

medieval theories of demonstration, and mental representation in medieval philosophy. For some important medieval authors in this area, see the entries on John Buridan, John Duns Scotus, Nicholas of Autrecourt, Saint Augustine, Saint Thomas, Walter Chatton, and William of Ockham.

For details on some important developments in medieval ethics, see the entries on medieval theories of conscience, medieval theories of practical reason, and the natural law tradition in ethics. For some of the major contributors to medieval ethics, see the articles John Duns Scotus, Peter Abelard, Peter of Spain Saint Anselm, Saint Augustine, Saint Thomas Aquinas and William of Ockham, elsewhere in this *Encyclopedia*. For some important figures in medieval political theory, see the articles Dante Alighieri, John Wyclif, John Wyclif's Political Philosophy and William of Ockham.

The above lists of topics and important figures should be regarded as only representative; they are far from exhaustive.

Bibliography

This bibliography includes only items cited in the body of the article, plus general resources relevant to the study of medieval philosophy. More specialized bibliographies relevant to particular topics and individuals may be found in other articles in this *Encyclopedia*. See the list of Related Entries below.

General Histories of Medieval Philosophy

Dronke, Peter (ed.), 1988, *A History of Twelfth-Century Western Philosophy*, Cambridge: Cambridge University Press.

Gersh, Stephen, 1986, *Middle Platonism and Neoplatonism: The Latin Tradition*, 2 volumes (Publications in Medieval Studies: Volume 23),

Saint Thomas Aquinas

First published Mon Jul 12, 1999; substantive revision Fri May 23, 2014

Thomas Aquinas (1225–1274) lived at a critical juncture of western culture when the arrival of the Aristotelian *corpus* in Latin translation reopened the question of the relation between faith and reason, calling into question the *modus vivendi* that had obtained for centuries. This crisis flared up just as universities were being founded. Thomas, after early studies at Montecassino, moved on to the University of Naples, where he met members of the new Dominican Order. It was at Naples too that Thomas had his first extended contact with the new learning. When he joined the Dominican Order he went north to study with Albertus Magnus, author of a paraphrase of the Aristotelian *corpus*. Thomas completed his studies at the University of Paris, which had been formed out of the monastic schools on the Left Bank and the cathedral school at Notre Dame. In two stints as a regent master Thomas defended the mendicant orders and, of greater historical importance, countered both the Averroistic interpretations of Aristotle and the Franciscan tendency to reject Greek philosophy. The result was a new *modus vivendi* between faith and philosophy which survived until the rise of the new physics. The Catholic Church has over the centuries regularly and consistently reaffirmed the central importance of Thomas's work, both theological and philosophical, for understanding its teachings concerning the Christian revelation, and his close textual commentaries on Aristotle represent a cultural resource which is now receiving increased recognition. The following account concentrates on Thomas the philosopher.

- 1. Life and Works
    - 1.1 *Vita Brevis*
    - 1.2 Education
    - 1.3 Writings

1. Life and Works

1.1 *Vita Brevis*

Thomas was born in 1225 at Roccasecca, a hilltop castle from which the great Benedictine abbey of Montecassino is not quite visible, midway between Rome and Naples. At the age of five, he was entered at Montecassino where his studies began. When the monastery became a battle site—not for the last time—Thomas was transferred by his family to the University of Naples. It was here that he came into contact with the "new" Aristotle and with the Order of Preachers or Dominicans, a recently founded mendicant order. He became a Dominican over the protests of his family and eventually went north to study, perhaps first briefly at Paris, then at Cologne with Albert the Great, whose interest in Aristotle strengthened Thomas's own predilections. Returned to Paris, he completed his studies, became a Master and for three years occupied one of the Dominican chairs in the Faculty of Theology. The next ten years were spent in various places in Italy, with the mobile papal court, at various Dominican houses, and eventually in Rome. From there he was called back to Paris to confront the controversy variously called Latin Averroism and Heterodox Aristotelianism. After this second three year stint, he was assigned to Naples. In 1274, on his way to the Council of Lyon, he fell ill and died on March 7 in the Cistercian abbey at Fossanova, which is perhaps twenty kilometers from Roccasecca.

1.2 Education

Little is known of Thomas's studies at Montecassino, but much is known of the shape that the monastic schools had taken. They were one of the principal conduits of the liberal arts tradition which stretches back to Cassiodorus Senator in the 6th century. The arts of the trivium (grammar, rhetoric, logic) and those of the quadrivium (arithmetic, geometry, music

and astronomy) were fragments preserved against the ruinous loss of classical knowledge. They constituted the secular education that complemented sacred doctrine as learned from the Bible. When Thomas transferred to Naples, his education in the arts continued. Here it would have been impressed upon him that the liberal arts were no longer adequate categories of secular learning: the new translations of Aristotle spelled the end of the liberal arts tradition, although the universities effected a transition rather than a breach.

Taking Thomas's alma mater Paris as reference point, the Faculty of Arts provided the point of entry to teen-aged boys. With the attainment of the Master of Arts at about the age of 20, one could go on to study in a higher faculty, law, medicine or theology. The theological program Thomas entered in Paris was a grueling one, with the master's typically attained in the early thirties. Extensive and progressively more intensive study of the scriptures, Old and New Testament, and of the summary of Christian doctrine called the *Sentences* which was compiled by the twelfth century Bishop of Paris, Peter Lombard. These close textual studies were complemented by public disputations and the even more unruly quodlibetal questions. With the faculty modeled more or less on the guilds, the student served a long apprenticeship, established his competence in stages, and eventually after a public examination was named a master and then gave his inaugural lecture.

1.3 Writings

Thomas's writings by and large show their provenance in his teaching duties. His commentary on the *Sentences* put the seal on his student days and many of his very early commentaries on Scripture have come down to us. But from the very beginning Thomas produces writings which would not have emerged from the usual tasks of the theological master. *On Being and Essence* and *The Principles of Nature* (the latter a very useful

summary for students of the principles that Aristotle develops in his *Physics*) date from his first stay at Paris, and unlike his commentaries on Boethius' *On the Trinity* and *De hebdomadibus*, are quite obviously philosophical works. Some of his disputed questions date from his first stint as regius master at Paris. When he returned to Italy his productivity increased. He finished the *Summa contra gentiles*, wrote various disputed questions and began the *Summa theologiae*. In 1268, at Rome, he began the work of commenting on Aristotle with *On the Soul*, and during the next five or six years commented on eleven more Aristotelian works (not all of these are complete). During this time he was caught up in magisterial duties of unusual scope and was writing such polemical works as *On the Eternity of the World* and *On There Being Only One Intellect*.

At Naples, he was given the task of elevating the status of the Dominican House of Studies. His writing continued until he had a mystical experience which made him think of all he had done as "mere straw." At the time of his death in 1274 he was under a cloud in Paris. 219 propositions were condemned in 1277 by a commission appointed by the Bishop of Paris, among them some tenets of Thomas. This was soon lifted, he was canonized and eventually was given the title of Common Doctor of the Church. But the subtle and delicate assimilation of Aristotle that characterized his work in both philosophy and theology did not survive his death, except in the Dominican Order, and has experienced ups and downs ever since.

2. Philosophy and Theology

Many contemporary philosophers are unsure how to read Thomas. He was in his primary and official profession a theologian. Nonetheless, we find among his writings works anyone would recognize as philosophical and the dozen commentaries on Aristotle increasingly enjoy the respect and interest of Aristotelian scholars. Even within theological works as such

there are extended discussions that are easily read as possessing a philosophical character. So his best known work, the *Summa theologiae,* is often cited by philosophers when Thomas's position on this or that issue is sought. How can a theological work provide grist for philosophical mills? How did Thomas distinguish between philosophy and theology?

Sometimes Thomas puts the difference this way: "… the believer and the philosopher consider creatures differently. The philosopher considers what belongs to their proper natures, while the believer considers only what is true of creatures insofar as they are related to God, for example, that they are created by God and are subject to him, and the like." (*Summa contra gentiles,* bk II, chap. 4) Since the philosopher too, according to Thomas, considers things as they relate to God, this statement does not put the difference in a formal light.

The first and major formal difference between philosophy and theology is found in their principles, that is, starting points. The presuppositions of the philosopher, that to which his discussions and arguments are ultimately driven back, are in the public domain, as it were. They are things that everyone in principle can know upon reflection; they are where disagreement between us must come to an end. These principles are not themselves the products of deductive proof—which does not of course mean that they are immune to rational analysis and inquiry—and thus they are said to be known by themselves (*per se*, as opposed to *per alia*). This is proportionately true of each of the sciences, where the most common principles just alluded to are in the background and the proper principles or starting points of the particular science function regionally as the common principles do across the whole terrain of thought and being. The fact that they are known *per se* does not imply that they are easily known to just anyone who considers them. A good deal of experience of the world and inquiry, not to mention native intelligence, and the ability to avoid

intellectual distraction, may be required for anyone in particular to actually apprehend their truth.

By contrast, the discourse of the theologian is ultimately driven back to starting points or principles that are held to be true on the basis of faith, that is, the truths that are authoritatively conveyed by Revelation as revealed by God. Some believers reflect on these truths and see other truths implied by them, spell out their interrelations and defend them against the accusation of being nonsense. Theological discourse and inquiry look like any other and is, needless to say, governed by the common principles of thought and being; but it is characterized formally by the fact that its arguments and analyses are taken to be truth-bearing only for one who accepts Scriptural revelation as true.

This provides a formal test for deciding whether a piece of discourse is philosophical or theological. If it relies only on truths anyone can be expected upon sufficient reflection to know about the world, and if it offers to lead to new truths on the basis of such truths, and only on that basis, then it is philosophical discourse. On the other hand, discourse whose cogency—not formal, but substantive—depends upon our accepting as true such claims as that there are three persons in one divine nature, that our salvation was effected by the sacrifice of Jesus, that Jesus is one person but two natures, one human, one divine, and the like, is theological discourse. Any appeal to an authoritative scriptural source as the necessary nexus in an argument is thereby other than philosophical discourse.

More will be said of this contrast later, but this is the essential difference Thomas recognizes between philosophy and theology. To conclude, consider a passage in which Thomas summarizes his position. He is confronting an objection to there being any need for theological discourse. Whatever can be the object of inquiry will qualify as a being of one sort or

another; but the philosophical disciplines seem to cover every kind of being, indeed there is even a part of it which Aristotle calls theology. So what need is there for discourse beyond philosophical discourse?

> … it should be noted that different ways of knowing (*ratio cognoscibilis*) give us different sciences. The astronomer and the natural philosopher both conclude that the earth is round, but the astronomer does this through a mathematical middle that is abstracted from matter, whereas the natural philosopher considers a middle lodged in matter. Thus there is nothing to prevent another science from treating in the light of divine revelation what the philosophical disciplines treat as knowable in the light of human reason. (*Summa theologiae*, Ia.1.1 ad 2)

For Thomas theological discourse begins with what God has revealed about Himself and His action in creating and redeeming the world. The world is understood in that light. Philosophical discourse begins with knowledge of the world. If it speaks of God what it says is conditioned by what is known of the world. But even given the distinction between the two, Aquinas suggests here that there are in fact elements of what God has revealed that are formally speaking philosophical and subject to philosophical discussion—though revealed they can be known and investigated without the precondition of faith. In other words, even something that is as a matter of fact revealed is subject to philosophical analysis, if religious faith is not necessary to know it and accept it as true. So it may happen that concerning certain subjects, as for example the nature of God, the nature of the human person, what is necessary for a human being to be good and to fulfill his or her destiny, and so on, there can be both a theological and a philosophical discussion of those subjects, providing for a fruitful engagement between the theological and the philosophical. For this reason, Thomas' theological works are very often

paradigms of that engagement between theological and philosophical reflection, and provide some of his very best philosophical reflection.

3. Christian Philosophy

It will be observed that the formal distinction between philosophical and theological discourse leaves untouched what has often been the mark of one who is at once a believer and a philosopher. It is not simply that he might on one occasion produce an argument that is philosophical and at another time one that is theological; his religious beliefs are clearly not put in escrow but are very much in evidence when he functions as a philosopher. Many of the questions that can be raised philosophically are such that the believer already holds a position on the answers to them from his religious faith. How then can he be thought to be ready to follow the argument whither it listeth, as an objector might put it? Furthermore, the inquiries in which the believer who philosophizes engages will often indicate his religious interests.

When such observations turn into objections, perhaps into the accusation that a believer cannot be a proper philosopher, there is often an unexamined notion of what a proper philosopher looks like. The proper philosopher may be thought to be someone—perhaps merely some mind—without antecedents or history who first comes to consciousness posing a philosophical question the answer to which is pursued without prejudice. But of course no human being and thus no philosopher is pure reason, mind alone, without previous history as he embarks on the task of philosophizing. One has necessarily knocked about in the world for a long time before he signs up for Philosophy 101. He has at hand or rattling around in his mind all kinds of ready responses to situations and questions. He very likely engaged in some kind of inquiry about whether or not to begin the formal study of philosophy in the first place. This may be acknowledged, but with the proviso that step one in the pursuit of

philosophy is to rid the mind of all such antecedents. They must be put in the dock, put in brackets, placed in doubt, regarded with suspicion. Only after appropriate epistemological cleansing is the mind equipped to make its first warranted knowledge claim. Knowledge thus becomes a deliverance of philosophy, a product of philosophizing. Outside of philosophy there is no knowledge.

The preceding paragraph has been meant to capture the salient note of much modern philosophy since Descartes. Philosophy is first of all a search for defensible knowledge claims, and for the method according to which it will be found. As opposed to what?

As opposed to the view of philosophy described in paragraph 2, Thomas understands philosophizing to depend upon antecedent knowledge, to proceed from it, and to be unintelligible unless, in its sophisticated modes, it can be traced back to the common truths known to all. But this tracing back will pass through very different terrains, depending on the upbringing, culture and other vagaries and accidents of a given person's experience. The pre-philosophical—I refer to the formal study of philosophy—outlook of the believer will be characterizable in a given way, a way suggested above. It is more difficult to characterize the pre-philosophical attitudes and beliefs out of which the non-believer philosophizes. Let us imagine that he holds in a more or less unexamined way that all events, including thinking, are physical events. If as a philosopher he should happen take up the question of the immortality of the soul, he is going to regard with suspicion those classical proofs which rely on an analysis of thinking as a non-physical process. The Christian, on the other hand, will be well-disposed towards efforts to prove the immortality of the human soul and will accordingly approach descriptions of thinking as non-physical sympathetically. He is unlikely to view with equanimity any claim that for human beings death is the utter end.

The importance of this is that a believer runs the risk of accepting bad proofs of the non-physical character of thinking and thus of the human soul. On the other hand, a committed physicalist may be too quick to accept a bad proof that thinking is just a physical process. He may be just as likely to run the risk of accepting bad proofs of the entirely physical character of thought as is the believer of the opposed claim. Such antecedent stances are often the reason why philosophical agreement is so hard to reach. Does it make it impossible? Do such considerations destroy any hope of philosophical objectivity on either side? Surely not in principle. Believers and non-believers should be able to agree on what counts as a good proof in a given area even if they expect different results from such a proof. Thinking either is or is not merely a physical process and antecedent expectations do not settle the question, however much they influence the pursuit of that objective resolution. But the important point is that antecedent dispositions and expectations are the common condition of philosophers, believers and unbelievers alike. Of course believers hold that they have an advantage here, since the antecedents that influence them are revealed truths, not just hearsay, received opinion, the *zeitgeist*, or prejudice. In addition they may be much more likely to be aware of and acknowledge those antecedents, insofar as they are explicitly held and inquired into.

4. Thomas and Aristotle

Given the distinction between philosophy and theology, one can then distinguish between philosophical and theological sources and influences in Aquinas' work. As a philosopher Thomas is emphatically Aristotelian. His interest in and perceptive understanding of the Stagyrite is present from his earliest years and did not await the period toward the end of his life when he wrote his close textual commentaries on Aristotle. When Thomas referred to Aristotle as the Philosopher, he was not merely

adopting a *façon de parler* of the time. He adopted Aristotle's analysis of physical objects, his view of place, time and motion, his proof of the prime mover, his cosmology. He made his own Aristotle's account of sense perception and intellectual knowledge. His moral philosophy is closely based on what he learned from Aristotle and in his commentary on the *Metaphysics* he provides a cogent and coherent account of what is going on in those difficult pages. Quite often deep insight into Thomas' philosophical thought can be gained from a close attention to the ways in which he comments upon and interpretively clarifies difficult passages in Aristotle that can be otherwise very obscure.

But to acknowledge the primary role of Aristotle in Thomas's philosophy is not to deny other philosophical influences. Augustine is a massively important presence. Boethius, Pseudo-Dionysius, and Proclus were conduits through which he learned Neo-platonism. There is nothing more obviously Aristotelian about Thomas than his assumption that there is something to be learned from any author and not only mistakes to be avoided. He definitely adopted many features from non-Aristotelian sources.

This has led some to suggest that what is called Thomistic philosophy is an eclectic hodgepodge, not a set of coherent disciplines. Others, struck by the prominence in Thomas of such Platonic notions as participation, have argued that his thought is fundamentally Platonic, not Aristotelian. Still others argue that that there is a radically original Thomistic philosophy which cannot be characterized by anything it shares with earlier thinkers, particularly Aristotle.

The recognition that Thomas is fundamentally an Aristotelian is not equivalent to the claim that Aristotle is the only influence on him. It is the claim that whatever Thomas takes on from other sources is held to be compatible with what he already holds in common with Aristotle. And, of

course, to draw attention to the sources of Thomas's philosophy is not to say that everything he holds philosophically can be parsed back into historical antecedents, or that he never disagrees with his sources, Aristotle in particular.

5. The Order of Philosophical Inquiry

Thomas takes "philosophy" to be an umbrella term which covers an ordered set of sciences. Philosophical thinking is characterized by its argumentative structure and a science is taken to be principally the discovery of the properties of kinds of things. But thinking is sometimes theoretical and sometimes practical. The practical use of the mind has as its object the guidance of some activity other than thinking—choosing in the case of moral action, some product in the case of art. The theoretical use of the mind has truth as its object. It seeks not to change the world but to understand it. Like Aristotle, Thomas holds that there is a plurality of both theoretical and practical sciences. Ethics, economics and politics are the practical sciences, while physics, mathematics and metaphysics are the theoretical sciences.

That is one way to lay out the various philosophical disciplines. But there is another that has to do with the appropriate order in which they should be studied. That order of learning is as follows: logic, mathematics, natural philosophy, moral philosophy, metaphysics. The primacy of logic in this order stems from the fact that we have to know what knowledge is so we will recognize that we have met its demands in a particular case. The study of mathematics comes early because little experience of the world is required to master it. But when we turn to knowledge of the physical world, there is an ever increasing dependence upon a wide and deep experience of things. Moral philosophy requires not only experience, but good upbringing and the ordering of the passions. Metaphysics or speculative wisdom is the culminating and defining goal of philosophical

inquiry: it is such knowledge as we can achieve of the divine, the first cause of all else.

Thomas commented on two logical works of Aristotle: *On Interpretation* (incomplete) and *Posterior Analytics*. On mathematics, there are only glancing allusions in Thomas's writings. Thomas describes logic as dealing with "second intentions," that is, with formal relations that attach to concepts expressive of the natures of existent things, first intentions. The concepts come to be in us as a result of our engagement with the sensible world. So it is important to stress that logic concerns formal relations between concepts. It is not concerned with the sort of developmental and causal relations studied within the the discipline of psychology. Thomas does not advocate a kind of proto-psychological account of logic as one might see in various 19th century accounts.

This means that logic rides piggy-back on direct knowledge of the world and thus incorporates the view that what is primary in our knowledge is the things of which we first form concepts. Mathematical entities are idealizations made by way of abstraction from our knowledge of sensible things. It is knowledge of sensible things which is primary and thus prior to the "order of learning" the philosophical sciences.

This epistemological primacy of knowledge of what we grasp by our senses is the basis for the primacy of the sensible in our language. Language is expressive of knowledge, and thus what is first and most easily knowable by us will be what our language first expresses. That is the rule. It is interesting to see its application in the development of the philosophy of nature.

6. Composition of Physical Objects

The concern of natural science is of course natural things, physical objects, which may be described as "what come to be as the result of a change and undergo change." The first task of natural philosophy, accordingly, is to define and analyze physical objects.

The first thing to notice about this is the assumption that we begin our study of the natural world not with the presumed ultimate alphabet with which macrocosmic things are spelled, but with a vague and comprehensive concept which encompasses whatever has come to be as the result of a change and undergoes change. The reader of Aquinas becomes familiar with this assumption. Thomas learned it from the beginning of Aristotle's *Physics*.

> The natural way of doing this is to start from the things which are more knowable and clear to us and to proceed towards those which are clearer and more knowable by nature; for the same things are not knowable relatively to us and knowable without qualification. So we must follow this method and advance from what is more obscure by nature, but clearer to us, towards what is more clear and more knowable by nature.

> Now what is to us plain and clear at first is rather confused masses, the elements and principles of which become known to us later by analysis. Thus we must advance from universals to particulars; it is a whole that is more knowable to sense-perception, and a universal is a kind of whole, comprehending many things within it, like parts. Much the same thing happens in the relation of the name to the formula. A name, e.g. 'Circle', means vaguely a sort of whole: its definition analyses this into particulars. Similarly a child begins by calling all men father and all women mother, but later on distinguishes each of them. (*Physics*, 1, 1.)

So, for example, in giving the 1st way of proving the existence of a god in the *Summa Theologiae*, Thomas will explicitly abide by this order when he says that we should begin with what is most manifest to us, namely, motion.

Thomas calls the movement from the more to the less general in a science the "order of determination" or specification of the subject matter. The first purchase on natural things is via "physical object" or "natural thing." The "order of demonstration" involves finding the properties of things as known through this general concept. Then, specifying the subject further, one seeks properties of things known through the less common concepts. For example, in plane geometry, one would begin with plane figure and discover what belongs to it as such. Then one would turn to, say, triangle and seek its properties, after which one would go on to scalene and isosceles. So one will, having determined what is true of things insofar as they are physical objects, go on to seek the properties of things which are physical objects of this kind or that, for example, living and non-living bodies.

Thomas emphasizes those passages in the Aristotelian natural writings which speak of the order of determination, that is, of what considerations come first and are presupposed to those that come later. In several places, Thomas takes great pains to array the Aristotelian natural writings according to this Aristotelian principle, most notably perhaps at the outset of his commentary on *Sense and sensibilia*. The *Physics* is the first step in the study of the natural world and exhibits the rule that what is first and most easily known by us are generalities. The language used to express knowledge of such generalities will have, as we shall emphasize, a long career in subsequent inquiries, both in natural philosophy and beyond. What is sometimes thought of as a technical vocabulary, perhaps even as Aristotelian jargon, is seen by Thomas Aquinas as exemplifying the rule that we name things as we know them and that we come to know more

difficult things after the easier things and extend the language used to speak of the easier, adjusting it to an ever expanding set of referents.

6.1 Matter and Form

Thomas points out that the characteristic features of *how* we know some subject should not in general be attributed to that subject as if elements of *what* we know of it. So, although natural things are first thought of and analyzed in the most general of terms, there are not any general physical objects, only particular ones. Thus, in seeking to discern what is true of anything that has come to be as a result of a change and is subject to change until it ceases to be, Aristotle had to begin with a particular example of change, one so obvious that we would not be distracted by any difficulties in accepting it as such. "A man becomes musical." Someone acquires a skill he did not previously have. Thomas pores over the analysis Aristotle provides of this instance of change and its product.

The change may be expressed in three ways:

1. A man becomes musical.
2. What is not-musical becomes musical.
3. A not-musical man becomes musical.

These are three different expressions of the same change and they all exhibit the form A becomes B. But change can also be expressed as From A, B comes to be. Could 1, 2 and 3 be restated in that second form? To say "From the not-musical the musical comes to be" and "From a not-musical man the musical comes to be" seem acceptable alternatives, but "From a man musical comes to be" would give us pause. Why? Unlike "A becomes B" the form "From A, B comes to be" suggests that in order for B to emerge, A must cease to be. This grounds the distinction between the grammatical subject of the sentence expressing a change and the subject of the change. The definition of the subject of the change is "that to which

the change is attributed and which survives the change." The grammatical subjects of 2 and 3 do not express the subject of the change. Only in 1 is the grammatical subject expressive of the subject of the change.

This makes clear that the different expressions of the change involve two things other than the subject of the change, namely, the characteristics of the subject before (not-musical) and after (musical) the change. These elements of the change get the names that stick from another example, whittling wood. The term for wood in Greek is *hyle* and the term for shape, the external contours of a thing, is *morphe*. In English, form, a synonym of shape, is used to express the characteristic that the subject acquires as the result of the change, e.g. musical. The characterization of the subject prior to the change as not having the form is called privation. Using this language as canonical, Aristotle speaks of the subject of the change as its *hyle* or matter, the character it gains as its *morphe* or form, and its prior lack of the form as its privation. Any change will involve these three elements: matter, form and privation. The product of a change involves two things: matter and form.

Change takes place in the various categories of quality, quantity, place, and so on. In all cases the terminology of matter, form, and privation comes to be used. So the terms applied in these different categories will be used analogously. The terms bind together similar but different kinds of change—a subject changing temperature is like a subject changing place or size.

6.2 Substantial Change

The analysis of change and the product of change begins with surface changes. Some enduring thing changes place or quality or quantity. But enduring things like men and trees and horses and the like have also come into being and are destined some day to cease to be. Such things are called

substances. It is a given that there are substances and that they come to be and pass away. The question is: Can the analysis of surface change be adjusted and applied to substantial change? What would its subject be? The subject of substantial change is known on an analogy with the subject of incidental or surface change. That is, if substances come to be as the result of a change, and if our analysis of change can apply, there must be a subject of the change. The subject of a surface or incidental change is a substance. The subject of a substantial change cannot be a substance; if it were, the result would be a modification of that substance, that is, an incidental change. But we are trying to understand how a substance itself comes into being as the result of a change. There must be a matter or subject but it cannot be matter in the sense of a substance. In order to signal this, we can call the matter *prime matter*, first matter. But it is important to recognize that this prime matter is not a substance, and does not exist apart from any particular substance. It is always the matter of some substance that exists.

When the discussion moves on from what may be said of all physical objects as such to an inquiry into living physical things, the analyses build upon those already completed. Thus, "soul" will be defined as the form of living bodies. The peculiar activities of living things will be grouped under headings like nutrition, growth, sense perception, knowing, and willing. Since a living thing sometimes manifests an instance of such activities and sometimes does not, they relate to it in the manner of the incidental forms of any physical object. But they are not incidental in the way that we might think of the shade of color of one's skin at any particular time, or the particular height or weight of an individual, since as activities the ability or power to engage in them proceeds from what the substance in question is. Thomas at times will call the powers through which they are achieved necessary accidents, using accident in a sense different from more recent philosophy. While the abilities need not be exercised at any particular time

or may be impeded from exercise by some condition, the substance nonetheless possesses them in principle as long as it exists.

The form such a subject takes on as the result of the change cannot be an incidental form like size or location or temperature. Substances do not become or cease to be substances as a result of changes in these incidental features. As the analysis of incidental change makes clear, the substance previously existed without the form it acquires in the change and it could lose it and still be itself. In a substantial change, the substance itself simply comes to be, or ceases to be. The form in a substantial change must be that which makes the substance to be what it is. Call it *substantial form*.

Here we see the semantic plasticity of the term 'matter'. Initially in the analysis of change, 'matter' refers to the substance that takes on or loses some incidental categorical modification of that substance. Then the term is extended by analogy to cover whatever is the subject of a change of substance. Recognizing the analogical extension, Thomas avoids "reifying" what he identifies as *principles* of change and nature. Although form and matter will be "reduced" in a substantial change to the category of substance as principles of substance, they should not be thought of as substances or quasi-substances in their own right. Socrates or Bucephalus is a substance strictly speaking. The forms and matter of Socrates and Bucephalus are not. They are substantial principles without being substances or quasi-substances in their own right.

So the point to notice about this analysis is that substantial change is spoken of on an analogy with incidental change. The analysis of incidental change is presupposed and regulative. Moreover, the language used to speak of the elements of incidental change are extended to substantial change and altered in meaning so as to avoid equivocation. The philosophical vocabulary arises out of analysis of what is most obvious to us and is then progressively extended to more and more things insofar as

the later is made known by appeal to the prior. We see that matter and form apply in an analogous way to the various kinds of incidental change and then to substantial change. The analysis of form and matter provides a rule for knowing and naming that will characterize Thomas's use of Latin in philosophy and in theology as well.

7. Perception and Thought

Focusing specifically upon perception—seeing, feeling, hearing, and the like—how can we best analyze it? In continuity with what has gone before, the questions are put in this form: How best to analyze coming to see, coming to feel, coming to hear, and the like? Seeing these on the analogy of change as already analyzed, we look for a subject, a privation, and a form. The sensing subject is the animal, but the proximate subjects to which they are attributed are the powers of sight, touch, hearing, and the like. An instance of seeing is describable as the power's moving from not seeing to seeing. Since the object of seeing is color, the change from not seeing to seeing issues in the power having the form of color.

Consider an ordinary physical change, a substance acquiring a color. Coming to see a color is not the same kind of physical change as a substance acquiring a color. To be sure, while there are physical changes involved in sensation—the organs are altered in the way physical bodies are—that is not the change involved in perception as such. Consider again that in feeling a warm or cold body the hand's own temperature is altered by the contact. But feeling cannot be just that, since any two physical bodies that come into contact undergo a similar alteration of temperature. But not all physical bodies feel the temperature. Feeling the temperature, becoming aware of it, is another sort of change, however much it involves a contemporaneous change in the organs of sense similar to ordinary physical change. Having the color or temperature in this further sense is thus made known and named by reference to physical change. The

fundamental difference between the two ways of acquiring a form is this: in a physical change of color, the change produces a new numerical instance of the color. In grasping or sensing a color, a numerically new instance of color does not result. And yet what was potentially visible becomes actually visible. There is actuality in the world where before there was only potentiality--an actuality of the seen color, and an actuality of color not in the mode of existence that color has in physical things.

We have here the basis for talk of immateriality in perception. If the acquiring of a form by matter in physical change results in a new instance of the form and this is not the case with perception, we can make the point that acquiring the form in sensation is not identical to the acquiring of the form by matter *in the primary sense*. Thus, we both want to speak of the subject of sensation on an analogy with physical change *and* to distinguish the former from the latter. This is done by speaking of the immaterial reception of a form. Nonetheless, the sense power is implemented in a physical organ, and thus matter for the change of form in sensation in an analogous sense. Because in sensation the sense organ is physically altered and the matter of sensation in this analogous sense, we can say that actual sensation is in some respects physical, and in another not.

It is important to pay attention again to the order of learning and naming, and what we are justified in saying at this point about the use of the words involved in describing this change. Specifically, the use of 'immaterial' is introduced simply to mark the inadequacy of any analysis of sensation confined solely to the physical terms that are fully adequate for analyzing ordinary physical change that does not involve sensation. 'Immaterial' means 'not-material'. But the mere applicability of such a negative term (what Aristotle calls a "negative infinite" term) does not justify us in thinking we have discovered a new property that would be referred to by the term 'immateriality'—it does not pick out and name a particular kind

of property—any more than the mere applicability of 'not-human' justifies us in thinking we have discovered a new particular kind of substance.

Such negative-infinite terms do not "cut at the joints" of reality.

Now, in his interpretation of Aristotle's *De anima* Thomas defends a view that was as contested in his own time as it is almost an orphan in our own. Among the tenets of so-called Latin Averroism was the view, first held by Averroes, that the move from perceptive acts to intellection is not one from a lower to a higher set of capacities or faculties of the human soul. Aristotle contrasts intellection with perception, and argues that the former does not employ a sense organ because it displays none of the characteristics of perception which does employ an organ. Thus insofar as sensation can be said to be in some respects material and in others immaterial, intellection is said to be completely immaterial. But on the Latin-Averroistic view, Aristotle is not thus referring to another capacity of the human soul, the intellect, but, rather, referring to a separate entity thanks to whose action human beings engage in what we call thinking. But the cause of this, the agent intellect, is not a faculty of the soul. (Aristotle had distinguished at least two intellects, a possible and an agent.) The proof for incorruptibility which results from an activity that does not employ a corporeal organ is therefore a statement about the incorruptibility of this separate entity, not a basis for arguing that each human soul is incorruptible because it has the capacity to perform incorporeal activities. The Latin-Averroists consequently denied that Aristotle taught personal immortality.

Given this consequence, Thomas's adoption of the opposite interpretation —viz. that the agent intellect is, like the possible intellect, a faculty of the human soul—may seem merely an interested desire to enlist Aristotle's support for a position in harmony with Christian belief. Thomas is frequently said to have baptized Aristotle, which seems to mean that he

fitted him to the Procrustean bed of Christian doctrine. Of course, the full Christian view is not simply that the soul survives death but that it will be reunited with body, and Thomas nowhere suggests that there is any intimation of this in Aristotle. Oddly enough, it is often friends of Thomas who suggest that he merely *used* Aristotle and was not chiefly concerned with what Aristotle might actually have intended.

However, this is an extraordinary approach to reading Thomas. It would be less of an accusation to say that he got a passage wrong than that he pretended it meant something he knew it did not. However, the important point is whether Thomas's reading is or is not supported by the text. When he commented on the *De anima*, he seems not to be concerned with the flare up in Paris over Latin Averroism. This is the basis for dating the commentary in 1268, before Thomas returned to Paris. The commentary, accordingly, cannot be read as though it were prompted by the controversy. Of course, some might still say that Thomas had long term interests in taming Aristotle to behave in a Christian way. On the contrary, as it happens, during the second Parisian period in the thick of the Latin-Averroist controversy, Thomas wrote an opusculum dedicated to the question: what did Aristotle actually teach? The work is called in the Latin, *De unitate intellectus contra averroistas*, *On there being only one intellect contra the Averroists*. This little work is absolutely essential for assessing the nature of Thomas's Aristotelianism. He provides us with an extended textual analysis to show that the rival interpretation cannot be sustained by the text and that the only coherent reading of the *De anima* must view the agent and possible intellects as faculties of the human soul. His interpretation may be right or wrong, but the matter must be decided on the basis of textual interpretation, not vague remarks about Thomas's intentions.

8. Body and Soul

Philosophers nowadays will want to know how this account of substance places Aquinas on the question of the relation of body and soul with respect to Dualism and Physicalism. Not easily. Aquinas maintains that the soul is capable of existing apart from the living body after the death of the body, because the soul is incorruptible. This might suggest that he is a kind of Substance Dualist, the soul being one substance and the body another, with the soul "interacting" as it were with the other substance, the body. However this picture fails to recognize the Aristotelian terms of the account that Aquinas provides of soul and body. Thomas knows and accepts Aristotle's assertion in *De anima II.1* that it is as pointless to ask whether soul and body are one as it is to ask whether the seal and the wax are one--they are.

The soul is indeed capable of existence apart from the body at death. This incorruptibility results from the actualities of understanding and willing that are not the actualities of any bodily organ, but of the human animal as such distinguished by the rational form. However, Thomas merely concludes from this fact that the soul is a "particular thing" and thus a subsistent after the death of the body. He argues that what belongs to the notion of "this particular thing" is only that it be a subsistent, and not that it be a substance complete in a nature. A subsistent is something with an operation of its own, existing either on its own or in another as an integral part, but not in the way either accidental or material forms exist in another. *Existing on its own* is not distinctive of substances alone. A chair is a particular thing, and thus a subsistent. But on Aquinas' account it is not a substance; it is rather an accidental unity of other subsistents which may or may not be substances. A hand has an operation distinctive of it as an integral part of a living body, an operation different from the operation of the stomach; it is a particular thing and also a subsistent. (*Summa Theologiae* Ia.75.2 ad1; also *Quaestiones Disputate de Anima* 2.) And yet being an integral and functional part of a substance, it does not have the complete nature of a substance.

A substance, on the other hand, is something that is both subsistent and complete in a nature—a nature being an intrinsic principle of movement and change in the subject. A human soul is a constitutive element of the nature of a human substance. It is the formal principle of a human substance. It is what is specified when we say what the substance is. But it is incomplete. What it is for a soul to be is to be the form of some substance. In that sense it is a principle of a substance, 'principle' being a technical term that refers back to the first entry, *arche*, in Aristotle's philosophical lexicon in the *Metaphysics*, as well as Thomas' commentary on it, and Thomas' *On the Principles of Nature*. As the principle of a nature, its nature is to be the formal element of a complete substance. Consequently, it doesn't have its own nature and is not a substance in its own right, even if it is capable of subsisting apart from the living body. It is because it is naturally incomplete as subsisting apart from the body that Thomas sees this state as unnatural for it, and an intimation of, but not an argument for, the resurrection of the body.

Question Ia.75 of the *Summa Theologiae* is the best place to look for Thomas' discussion of the subsistent reality of the human soul, although the *Quaestiones Disputatae De Anima* and the commentary on Aristotle's *De anima* provide important supplementary material to that discussion. Thomas begins 75 by pointing out that his concern is the concern of a theologian, and that the theologian is concerned with human nature primarily in relation to the soul. He is concerned with the body only in its relation to the soul. The body of the question is filled with philosophical argument, and yet its order and point is theological. That theological order and point, however, can lead to certain philosophical distortions concerning the soul if one isn't careful. So Thomas is very careful.

Considered as a substantial form of a material body, the soul exists in a living being as the substantial form of an animal. Here it is important to clarify. 'Immaterial' can be said in two ways of forms. In the first way, any

form as such is *immaterial* because it is not a material principle. It is distinguished as a principle of actuality in a being from the material principle which is a principle of potentiality and change in corporeal beings. In that sense, any substantial form whatsoever will be immaterial, including the substantial form of an oak tree or the substantial form of a dog. And so also is the substantial form of the human *immaterial* in that sense. Aquinas is explicit about this when he proves that the human soul is immaterial in *Summa Theologiae* Ia.75.5. It is immaterial in just the way in which any form whatsoever is immaterial. But in the second way, 'immaterial' is said of subsistent forms—forms that subsist without matter like angels or spiritual substances in general.

In 75.1 Thomas had argued against the ancient materialists, that the soul is not a body; it is incorporeal. In 75.2 he proved that the human soul is a subsistent because it has an activity that pertains to it without the use of a corporeal organ, namely, the activity of understanding in intellect. But then immediately in 75.3 he proved that the souls of other animals are not subsistent, because they do not have an operation that does not employ a corporeal organ. The souls of other animals are incorporeal in the sense of 75.1, but they are not subsistent as in 75.2 In 75.4, Thomas proves that the soul is not the man. Socrates, the man, has vital activities that are the activities of a living animal, like sensation, nutrition, reproduction, and so on, activities that are not distinctive activities of the soul itself as intellect is in the human case. Since these are activities of Socrates and not activities of the soul, Socrates and the soul are not identical. And so Socrates, if anything, is a living animal just like the other animals. Tacitly this leaves open the possibility that there might be an animal soul for Socrates that is not identical to the intellectual soul, and as shown in 75.3 that this animal soul of Socrates would not be subsistent. This possibility of two souls in Socrates, an animal soul and an intellectual soul will only be excluded later in question 76. In 75.5 Thomas proves that the human intellectual soul is immaterial just like the souls of other animals. But in

conjunction with the result of 75.2, now we have a soul that is an immaterial subsistent, where in conjunction with 75.3 the souls of other animals are not immaterial subsistents.

In 75.6, relying upon all that has gone before, Thomas argues that the human soul is a subsistent that is incorporeal, and thus does not cease to exist as a result of the death of the body. This result shows the soul to be a subsistent form that can exist without out matter. And so it is now seen to be an immaterial subsistent in the second sense described above, not just the first sense. Now 'immaterial' characterizes its mode of existence, not just the negative fact that it is immaterial like all other forms are immaterial.

So the difference between the human intellectual soul and the souls of other animals is that while both are immaterial in the first sense, the sense of not being material principles, the intellectual soul is an immaterial subsistent in the second sense while the souls of other animals are not immaterial subsistents. And it is the second sense of 'immaterial' that gives us a key for understanding what Thomas means by a "material form," particularly a material substantial form. A material form is a form that is not an immaterial subsistent; it exists either as an accident in a corporeal subject or as a substantial form in a corporeal subject, and does not subsist. So the substantial forms of bodies, particularly the souls of living bodies, are in general material forms with the exception of the intellectual soul. The souls of other animals are immaterial in the first sense and material with regard to the second sense, while the human soul is both immaterial in the first sense and immaterial in the second sense.

Confirmation of this distinction of senses of 'immaterial' comes when in the very last article of the question, 75.7, Thomas asked whether the human intellectual soul is an angel. In 75.6 Thomas used the result of 75.5 and 75.2 to prove that the human soul is incorruptible, where something is

corruptible if it can cease to exist through corruption, that is, by the separation of form from matter as we see it in the death of living things. The souls of other animals are not *directly* generated and do not *directly* corrupt. It is the living animal that corrupts. But their souls can be said to corrupt *with* the animal. (*Quaestiones Disputatae De Anima* 2) However, the human soul, because it is a subsistent immaterial form, does not corrupt *with* the death of the human being. So when all these results are put together the intellectual soul is an incorporeal, immaterial, incorruptible subsistent, an immaterial form in the second sense, which looks an awful lot like an angel, since angels are also incorporeal, immaterial, incorruptible subsistents, and immaterial forms in the second sense. This is the potentially distorting view of the theologian--to see the human intellectual soul as something like an angel, and we are reminded of Shakespeare, "how like an angel." Notice, however, that the potential distortion is premised on the soundness of the philosophical arguments that have been employed throughout the question by the theologian, driven by his primary interest in the soul.

In 75.7, Thomas argues that the intelletual soul is not of the same species as an angel, because it is a substantial form of an animal. Angels are complete in their natures as incorporeal, immaterial, incorruptible subsistent forms—they are thus substances properly speaking. But Thomas had insisted all along that the soul is incomplete in its nature, even as it is an incorporeal, immaterial, incorruptible subsistent form—it is not a substance properly speaking. Still, the soul can be called *substance* by analogy, insofar as it is the formal principle of a substance. In English it might be better to call it "substantial" rather than "substance." And in that regard, it cannot be considered as forming the basis for a kind of substance dualism in Thomas.

The argument of 75.7 leads naturally to the subject of the next major question, Ia.76, on the union of soul and body. We've already seen that

Thomas, following Aristotle, thinks asking questions about the union of soul and body makes little sense for the philosopher. But because of the potentially distorting view of the theologian, the latter in a sense is forced to do so; the theologian has to ask philosophical questions the philosopher need not ask, in order to avoid a distorted view of the soul. So in question 76 Thomas argues for the complete unity of soul with body against various alternative positions to be found among his contemporary theological interlocutors. Thus question 75, proceeding as it does from the theological perspective, gives rise to philosophical *aporiae* to be solved in question 76. And just as it was the theologian's use of philosophical arguments in 75 that threatened a distorted view, it is the theologian's use of philosophical arguments in 76 that solves the *aporiae*, and avoids the distortion. Apart from anything else Thomas does in the two questions, taken together they provide an exemplar of the use of philosophy within theology, not just to advance certain theological positions but to assist the theologian in avoiding error given the exclusivity of his theological perspective. Thomas fulfills what he himself had said is one of the roles of philosophy within theology in the first question of the *Summa*.

There are at least three important results of Ia.75-76 for thinking about human nature. In the first place, in 76.3-4 Thomas argues against the pluralist position on the Plurality of Substantial Forms. It might be tempting to think of the human substantial form as a kind of layering of quasi substantial forms or as composed out of them. One substantial form for the corporeality of the body, perhaps one to account for the vegetative activities of the human being, yet another for the animal activities, and then a final one for the intellectual activities of the human being. Recall that 75.4, in arguing that Socrates is not identical to the intellectual soul, tacitly leaves open the possibility that Socrates might have a soul as principle for these vital animal activities in addition to the intellectual soul with which he is not identical. However, Thomas decisively rejects this plurality on the basis of the manifest unity of the human being in his acts.

If there were multiple substantial forms there would be no unity to being human—multiple substantial forms implies multiple substances and multiple beings. And yet the human being is one, a single substantial unity manifested in his or her acts. Here Thomas is relying upon the substantial unity that is obvious to the philosopher to reject a kind of substance plurality, not just soul-body dualism.

In particular he relies upon the fact that it is Socrates himself who engages in intellectual activity. Again, in 75.4 he had rejected the view that Socrates is identical with his soul because Socrates engages in vital animal activities that do not belong to the soul as such. However, what he did not claim in 75.4 is that the activity of intellect that characterizes the soul is not one of Socrates activities. In fact, now in 76 he claims it is Socrates' activity. Socrates has vital activities that do not belong to the soul alone, and yet the activity that belongs to the soul alone, understanding, is one of Socrates' activities. But the soul is the principle of activity in living things. Thus the animal soul (and for similar reasons the vegetative soul) is identical in Socrates with the rational soul. There is no plurality of substantial forms because of the unity of Socrates' activities, including both animal activities and reason. Neither is the human soul composed of any quasi-substantial forms.

This is the second striking result of 76. Socrates and his soul, while not being identical, are subjects of the same activity—not subjects of the same *type* of activity, but subjects of the same *token* instance of an activity. In 75, the soul as a subsistent with its own operation of understanding was said to be the subject of existence (*esse*) *per se*. In the case of other animals it is the animal itself, the living substance, that is the subject of the act of existence, and both soul and body have existence through the substance. Here in the human case, the soul is said to be the subject of the act of existence because it has its own operation. Of course, Socrates is a substance with operations that pertain to him, animal activities, but also

the operation of intellect; it is Socrates who thinks in virtue of his intellect. So he too is the subject of the act of existence. And yet the operation in virtue of which the soul is the subject of the act of existence, intellectual activity, is the operation in virtue of which Socrates is the subject of the act of existence, again, not the same type of operation but the same token of operation. So Socrates and his soul have the same act of existence. The principle for drawing this latter conclusion is that the operation of a subject follows from the act of existing of that subject, as the actuality of a power follows from the actuality of the being. (*Quaestiones Disputatae De Anima* 2)

So Socrates, as a living animal substance, is not identical to his soul. *Anima mea non est ego* Thomas asserts in his Commentary on St. Paul's First Letter to the Corinthians. "I am not my soul." And yet Socrates and his soul share as subjects the very same act of existence. It is because Socrates' soul's act of existence is Socrates' act of existence that the soul's intellectual operation is Socrates' intellectual operation. It is also because of this sharing in the act of existence, that the soul can be the substantial form of the living human animal. Because the soul is a substantial form, it is not complete in its nature, and cannot be a spiritual substance like an angel, properly speaking. Thus the soul receives its act of existence *as* the soul of a human being, and cannot pre-exist the human being whose soul it is. And yet, as Thomas argued in 75.6, the soul is not subject to the corruptibility of the human being who is subject to death, and can thus survive the death of the human being.

The third significant result is that the soul is not composed from its powers as if a unified collection of them. Thomas will often speak of the "parts of the soul," the appetitive part or the rational part, for example. However, this way of speaking is for the purposes of classifying the powers. It does not signal actual ontological parts of the soul. As the first act of a body, the soul is, like all act, ontologically simple, undivided, and un-composed.

The powers "flow" from the essence of the soul as necessary accidents of the substance. And Thomas tells us they are formally related to the soul as their principle in what Aristotle calls in the *Posterior Analytics* the second mode of *per se* predication—that mode in which the subject of a predication enters into the definition of the predicate, were one to define the predicate. From this it follows that if the human soul is incorruptible, the powers of Socrates that are powers of corporeal organs cease to exist with the death of Socrates. And yet the power of intellect as a power of the soul without a corporeal organ remains incorruptible with the human soul. However, Thomas is clear in denying that only the intellect survives the death of the human; one cannot have a free floating incorruptible power in existence without the subject of the power in existence. The soul is other than the powers, and thus the intellectual power is incorruptible as a power of an incorruptible soul as its subject, and even the other powers remain "in" the soul as "in their principle."

All of this emphasis upon the unity of the human being comes out clearly in Aquinas' understanding of the mode of human activity as acting knowingly and willingly. Such acting knowingly and willingly is expressed as the rational activity of an animal, that is, as animal activity distinguished formally as rational. Rationality is the distinctive form that intelligence takes in human beings as animals. Rationality involves the back and forth of argument moving from one thing known to another, and advancing in knowledge by such movement. Thus, for Thomas, while angels and God can be said to be intelligent, they are not rational.

This movement in understanding is necessary for human beings because as animals they only ever have a partial grasp of the natures of things, insofar as their knowledge depends upon always incomplete and partial sensible experience of the world. But it is sense experience, as well as the self movement that springs from it, that places human beings within the genus animal. So human understanding and willing is intrinsically bound up with

the sensate activity of an animal; as a result, *rational* is the form that it takes in that animal. Reason does not cause eating as something separate from it, and as an efficient cause; on the contrary, human eating is not adequately described formally unless it is described as rational eating. To fail to eat rationally is not a failure in its cause, but in the eating itself. And the human animal is not adequately described except as a rational animal, rational providing not another substance or expression of a fissure between soul or mind and body, but the fully adequate description of the human substance. Reason does not distinguish us from animals; it distinguishes us as animals. So according to Aquinas, while it is true that the activities of intellect and will are not the actualities of any physical organs, they are nonetheless the activities of the living human animal. It is Socrates the animal who knows and wills, not his mind interacting with his body.

Another consequence of this insistence on Aquinas' part is that it is inadequate and inaccurate to speak of activities we share in common with other kinds of creatures. To be sure, there are descriptions that apply equally to what we do and what other animals do, for example the description "eating" or the description "reproducing." But these are generic descriptions that do not adequately capture the human act as opposed to the act of a horse or dog, until the human acts are specified formally as rational. So the goods that are the objects of human powers are not specified adequately by such generic descriptions as pursuing eating, reproducing, friendship, etc., as if human beings and other animals pursue the same goods, only humans bring reason to bear upon those identical type goods.

All of this might lead one to think then that, not being a dualist, Aquinas must be a physicalist, there being only two broad possible positions. Now, the difficulties of providing an adequate account of just what Physicalism is are well known. But suppose we take a minimal characterization of Physicalism as involving the claim that there is some privileged physical

science or set of physical sciences, using the term 'physical' merely nominally and sociologically as we use it of certain sciences today, that ideally will provide a fully adequate account of all that exists and the fundamental characteristics of reality. Then Aquinas cannot be understood to be a physicalist, since the result of his analysis of perception and thought was to say that these activities are "immaterial," which was to say, not adequately captured by the kinds of physical descriptions that do adequately account for much of the being and change we observe in the world. There are actually many variations on Dualism and Physicalism in play in recent philosophy. However, the difficulty of placing Aquinas in the broad outlines of that setting ought now to be clear. And without an actual demonstration that Aquinas' view is incoherent, one lasting contribution of his thought is to show that the supposed exclusive disjunction between Physicalism and Dualism is inadequate. He poses to us a challenge to think more broadly and deeply about human existence than such an easy dichotomy allows.

9. Human Identity and Immortality

To be immortal is not to be subject to death. Living corporeal substances are subject to death through the corruption of their substantial unity—not so much the separation of soul from body, but the dissolution of the soul as the substantial form of the body. This dissolution of the soul is brought about by destructive natural causes acting upon the living body. Living things themselves have various capacities to preserve themselves in existence against the ravages of the natural world around them; that is, in part, what it is for them to live—to sustain their existence in and through their own natural activities. And yet nature teaches us that corruptible things inevitably corrupt.

What of Socrates? Socrates is an animal. Thomas is unambiguous about this fact in 75.4 when he argues that Socrates is not identical to his soul

because of his animal acts. But, if Socrates is an animal he should be as subject to corruption as is any corporeal substance, and as subject to death as any animal. Here it is important to make an initial distinction. As we've seen, living things act to preserve their existence through their vital activities, and succeed in doing so for a time, even if they eventually succumb to the reaper. So we may say they are naturally subject to death because of their composed corporeal natures. And yet, it does not follow that they must corrupt and die; by and large their lives consist in preventing the corruption to which they are subject. This fact alone shows that there are causes with the power to fend off death, even if not to fend it off permanently. However, it is then at least possible that some other cause, a cause with much greater power than the natural causes of living things possess, could fend off death for them without end and preserve them alive without end. The obvious candidate for this cause is God by miraculous intervention; if living things have limited power to fend off their own deaths, presumably God has unlimited power to do so for them. So what is corruptible by nature may not in fact corrupt. While animals are naturally subject to death, they could be supernaturally immortal.

So also even if human beings are naturally subject to death, it may well be within the power of God to keep them from dying by a preternatural gift. This condition of having been given a preternatural gift preserving them from death would be the condition of the first human beings in the biblical account of Eden, the preternatural gift lost by Original Sin through which death entered into the world, however else one understands those data of revelation. But philosophically we can say no more of them than that human beings are naturally subject to death but need not die.

However, the world we live in is not an Edenic paradise into which death has not entered. Living things die. Human beings die. Socrates died. On the other hand, according to Thomas, Socrates' soul is incorruptible where the souls of other animals are not. It is not even naturally subject to death

by corruption. Is there a possibility for immortality, particularly personal immortality, here in the incorruptibility of Socrates' soul?

One might be tempted to say yes. One might say that in the first place the incorruptible soul of Socrates looks like a person in the current sense of that term. It is a thinking or conscious thing, since it is clearly a thing at least in the sense of a subsistent, and it has the power of intellect, even if it has no other conscious cognitive powers of the animal for which it formally was a soul. What "we" mean by person is a "thinking or conscious thing." So Socrates' soul is a person. What person? Well Socrates was a person in that very sense as well, although he had more conscious cognitive capacities than does his soul after death. It seems incongruous to suggest that we have two persons—Socrates and Socrates' soul. After all that would seem to strike against the unity Thomas was at pains to maintain. While Socrates was alive, were these two persons present? No. There is but one person, and it is Socrates. But then upon the death of Socrates, what happens? Does the person who is Socrates cease to exist, and a new person that is Socrates' soul come to exist? But, it seems much easier and simpler to say that upon Socrates' death the person that was Socrates survives as Socrates's soul. Before death Socrates was composed of a soul and a body. After death he is composed simply of a soul. If we hold that position then, because of the incorruptibility of the soul, while the animal that Socrates was dies, the soul that Socrates becomes survives, and thus Socrates himself is immortal, and not subject to death, not subject to death even by nature as the animal is. Socrates is simply immortal. On this proposed account when Thomas wrote "*anima mea non est ego*," he was merely speaking of this life. He should then have added "*anima mea ero ego*," "I will be my soul."

However as an interpretation of Thomas this approach will suffer several severe difficulties. First, on its own terms it is hard to avoid the conclusion that before Socrates' death, there are two persons present. The argument of

75.2 was not that the intellectual soul *becomes* a particular thing and subsistent upon the death of the human being. It was that the intellectual soul as such is a particular thing and subsistent, and that includes while it is the soul of a living thing. So if we are going to take the recent minimalist account of person that the term expresses in this proposed interpretation, *a thinking or conscious thing*, then we have the person that is the particular and subsistent thing that is the soul before the death of Socrates. But Thomas thinks Socrates thinks, and is thus a thinking thing. So we also have the person Socrates. Is the person that is the soul identical to the person that is Socrates? It seems not, given the argument of 75.4 that Socrates is not identical to his soul. So this interpretation suggests that even if after death there is only one person, Socrates, before death there are two persons, Socrates and Socrates' soul.

In the second place, this interpretation explicitly relies upon an equivocation on the term 'person'. Thomas accepts from Boethius the definition of a person as *an individual substance of a rational nature*. (*Summa Theologiae* IIIa.2.2) But Thomas insists that the soul is incomplete in its nature. It does not *have* a nature, but is one of the principles of a corporeal nature along with matter. It can only be said to "have" a nature improperly and by analogy. And when we do so speak, what is meant is that its nature is to be the substantial form of an animal. Again, that is why it is not an angel. So strictly speaking, the human soul, even as a subsistent, is not and cannot be a person, unless one equivocates on the term, and in so doing abandons the Philosophy of Nature and Metaphysics within which Thomas thinks.

In the third place, this interpretation would make hay of Thomas' argument in 75.4 that Socrates is not identical to his soul. There Thomas relied upon the vital activities of Socrates to make that argument--Socrates has vital activities that the soul does not possess as a subject or subsistent. But they are Socrates' activities as agent just as much as is the operation of intellect.

The powers that those activities manifest are powers of Socrates in just the way the power of intellect is. Again this resolves into the "nature" of the soul as substantial form. All of Socrates' powers "flow from" the soul as their formal principle, even as one of them, intellect, also has the soul as its subject along with Socrates. So if one were to ask which of the powers might be thought to be not quite Socrates' power in the full sense, one ought to opt for the intellect, not the vital powers of the living body, since it seems that intellect belongs to something other that Socrates and is at best shared with Socrates. But then why would Socrates become identical to the subject in virtue of a power that is not quite his, rather than cease to be with the powers that are properly his? Such questions, and the answers one might give to them, are again senseless if we situate what Thomas thinks back in what he wrote. The reason that intellectual power is no less Socrates' power than it is the soul's is because the act of being of Socrates is the act of being of his soul. It is a mistake to think that because Socrates is not identical to his soul, his soul forms some other being with which he would share some power. Again, this has to do with the soul being his substantial form.

In the fourth place, this interpretation would suggest, in Thomas' terms, that the body with its powers is *per accidens* related to Socrates' being. If Socrates is a substance, and the body is *per accidens* to his being, then the body is *per accidens* to his substance. In which case he is not a corporeal substance or animal at all, even in this life. The interpretation seems to return to giving the appearance that the intellectual soul is a kind of angel, only now adding that this angel is Socrates for a time associated with bodily powers. But recall Thomas' rejection of the Plurality of Substantial Forms position. His own account of the soul is that the animal powers of the soul are as much powers of the human soul as is the intellectual power —they are all powers in the second mode of *per se* predication. In that respect they are all alike, and the human soul is thus *per se* the substantial form of a living body, not *per accidens*, and the person Socrates is that

living body. When that living body ceases to exist through death, so also does the person who is Socratres. "*Anima mea non est ego*" simply.

Finally, Thomas clearly understands and accepts the implications of his view that Socrates is the living animal, namely, that the continued existence of the human soul after death is not sufficient for the continued existence of the human person. If the living animal no longer exists after death, then neither does Socrates. If the living animal is not immortal, then neither is Socrates. Consider these objections that Thomas himself considers. There is no resurrection of the body; only the souls of Abraham, Isaac, and Jacob live after death. Thomas writes in response that the soul of Abraham is not Abraham, and the life of Abraham's soul is not sufficient for the life of Abraham. The whole composite of Abraham's soul and body must live for Abraham to live. Thus if only Abraham's soul lives after death, Abraham does not. (*Summa Theologiae* 75.1 ad 2) Similarly when in the Commentary on First Corinthians Thomas writes "*anima mea non est ego*," it is precisely to emphasize the importance of the resurrection of the body for human salvation. He writes that "hence allowing that the soul gains salvation in another life, neither I nor any other human being does," (*unde licet anima consequatur salutem in alia vita, non tamen ego vel quilibet homo*.) because "anima mea non est ego." The importance of the theological doctrine of Christ's Resurrection is that it affirms that it is I who am ultimately saved through my resurrection from the dead, which involves me coming to exist when my body rises from the dead.

So while Socrates was not in this life actually immortal (he did die after all), he may in the resurrection live again and be made immortal by God.

Of course Thomas does not think that the resurrection of the body is demonstrable in philosophy. For him it is a revealed truth, not one of the *praeambula fidei*. Nonetheless he thinks it is rationally fitting for God to

bring it about, since otherwise there would be these very odd beings in existence for all time, incorporeal, immaterial, incorruptible subsistent forms the "nature" of which is to be the substantial forms of living bodies. Earlier we saw how Thomas' use of philosophical analysis helped to avoid the potentially distorting view of the theologian upon the nature of the soul. Here, we see how a revealed truth helps the philosopher avoid an equally distorting philosophical account of the soul and personal identity that would skew the philosophical books toward a personal human immortality without having to live as a human animal.

10. Beyond Physics

When Aristotle rejected the Platonic Ideas or Forms, accepting some of the arguments against them that Plato himself had devised in the *Parmenides*, he did not thereby reject the notion that the telos of philosophical enquiry is a wisdom which turns on what man can know of God. The magnificent panorama provided at the beginning of the *Metaphysics* as gloss on the claim that all men naturally desire to know rises to and culminates in the conception of wisdom as knowledge of all things in their ultimate or first causes.

For much of the twentieth century, Aristotelian studies had been conducted under the influence of Werner Jaeger's (1934) evolutionary hypothesis. On this view, Aristotle began as an ardent Platonist for whom the really real lay beyond sensible reality. With maturity, however, came the sober Macedonian empiricism which trained its attention on the things of this world and eschewed all efforts to transcend it. As for the *Metaphysics*, Jaeger saw it as an amalgam of both theories. The passage just alluded to at the beginning of the work is ascribed to the Platonic phase. Other passages have a far more modest understanding of the range and point of a science over and above natural philosophy and mathematics. *Platonice loquendo*, there are entities which exist separately

from sensible things and they constitute the object of the higher science. The more sober view finds a role for a science beyond natural philosophy and mathematics, but it will deal with things those particular sciences leave unattended, e.g. defense of the first principle of reasoning. But these tasks do not call for, and do not imply, a range of beings over and above sensible things.

Jaeger (1934) found both these conceptions of metaphysics juxtaposed in a crucial passage of Book Six.

> One might indeed raise the question whether first philosophy is universal, or deals with one genus, i.e. some one kind of being; for not even the mathematical sciences are all alike in this respect,— geometry and astronomy deal with a certain particular kind of thing, while universal mathematics applies alike to all. We answer that if there is no substance other than those which are formed by nature, natural science will be the first science; but if there is an immovable substance, the science of this must be prior and must be first philosophy, and universal in this way, because it is first. And it will belong to this to consider being *qua* being—both what it is and the attributes which belong to it *qua* being. (1025a24–33)

Jaeger invites us to see here a monument to a lost hope and an abiding reluctance to bid it a definitive farewell. Aristotle mentions the possibility of an immovable substance, something existing apart from the natural realm. Without such a separate substance, natural philosophy will be first philosophy. If there is such a substance, it will be a kind of being different from material being. The science that studies it will bear on a certain kind of being, immovable substance, immaterial being, not on being as being. It will be a special, not a universal, science. Jaeger sees Aristotle seeking to glue on to the special science the tasks that belong to a universal science, to make a theology into an ontology.

Jaeger's hypothesis dominated interpretations of the *Metaphysics* until very recently. Giovanni Reale's book (1961) had to await English translation before it could have any impact in English circles of interpretation. By that time, people were turning from Jaeger's account and toward a more direct reading of Aristotle. When we reconsider Thomas as a commentator on the *Metaphysics*, it becomes clear that his reading is in stark opposition to Jaeger's claims.

But let us first lay out Thomas's view of metaphysics. His question is Aristotle's: is there any science beyond natural science and mathematics? If to be and to be material are identical, then the science of being as being will be identical with the science of material being. That is what Aristotle rejects in the passage just quoted. It is in the course of doing natural philosophy that one gains certain knowledge that not everything that is is material. At the end of the *Physics*, Aristotle argues from the nature of moved movers that they require a first unmoved mover. If successful, this proof establishes that there is a first mover of all moved movers which is not itself material. Furthermore, the discussion of intellect in *On the Soul III*, to which we alluded in the preceding paragraph, points beyond the material world. If the activity of intellect provides a basis for saying that, while the human soul is the substantial form of the body, it can exist apart from the body, that is, survive death, it is an immaterial existent. The Prime Mover and the immortal souls of human beings entail that to be and to be material are not identical. Since these are acquisitions at the limit of natural philosophy, they represent possible objects of inquiry in their own right. This is pre-eminently the case with the Prime Mover. It seems inevitable that there should be a discipline whose principal aim is to know more about the divine. How can it be described?

Thomas discusses early the way theoretical sciences are distinguished from one another in the course of his exposition of the tractate of Boethius *On the Trinity*. The text speaks of three kinds of theoretical science,

physics, mathematics and theology, and Thomas invokes the methodology of the *Posterior Analytics*. A *scientia* is constituted by a demonstrative syllogism. From a formal point of view, a conclusion follows necessarily from the premises in a well-formed syllogism. Still the conclusion may state a merely contingent truth. What is needed in a demonstrative syllogism is not just the necessity of the consequence but a necessary consequent, and this requires that the premises express necessary truths. That which is necessary cannot be otherwise than as it is; it cannot change. Science thus requires that it bear on immobile things. There is another requirement of the object of speculative or theoretical knowledge which stems from intellection. The activity of the mind, as has been mentioned, is not a material event; it is immaterial. Since it is the mind that knows, science is a mode of its knowing, and will share its nature. Thomas thus states two essential characteristics of the object of speculation, the *speculabile*: it must be removed both from matter and from motion. If that is the case then insofar as there are formally different ways in which *speculabilia* can be removed from matter and motion, there will be formally different speculative sciences.

By this analysis, Thomas has provided the necessary background for understanding the text of Boethius but also more importantly that of Aristotle as it is developed in the chapter from which Werner Jaeger quoted in order to display the failure of the Aristotelian project. "Now we must not fail to notice the nature of the essence and of its formula, for, without this, inquiry is but idle. Of things defined, i.e. of essences, some are like snub, and some like concave. And these differ because snub is bound up with matter (for what is snub is a concave *nose*), while concavity is independent of perceptible matter." (1025a28–32) The objects of natural philosophy are defined like 'snub' and the objects of mathematics like 'concave'. This makes it clear that the way in which natural things are separated from sensible matter is the way in which the definition common to many things abstracts from the singular characteristics of each. But it is

the matter as singular that is the principle of change in things, so the common definition has the requisite necessity for science. This or that man comes to be, but what-it-is-to-be-a-man does not come to be or pass away.

Mathematical things, on the analogy of 'concave', do not have sensible matter in their definitions. Lines, points, numbers, triangles—these do not have sensible qualities whether stated universally or singularly. The fact that we define mathematicals without sensible matter does not commit us to the view that mathematicals actually exist apart from sensible matter.

In the commentary on Boethius to which reference has been made, Thomas has early on recalled another fundamental aspect of Aristotle's thought. The objects of thought are either simple or complex, where complex means that one thing is affirmed or denied of another. Knowledge of simples is expressed in a definition, that of the complex in a proposition. Thinking of human nature without thinking of singular characters of this man or that is a matter of definition, not of assertion, as if one were denying that human nature is found in singular matter. So too defining mathematicals without sensible matter is not tantamount to the judgment that mathematicals exist apart from sensible matter. These are both instances of abstraction, where abstraction means to think apart what does not exist apart. Thus it is that the question of metaphysics turns on what Thomas calls *separatio*. To separate differs from abstraction in this that separation is expressed in a negative judgment, an asserted proposition: this is not that, that *this* exists apart from *that*. The relevant separation for metaphysics is the negative judgment that to be and to be material are not the same. That is, there are things which exist apart from matter and motion—not just are defined without, but exist without matter and motion.

What then is the subject of metaphysics? "Subject" here means the subject of the conclusion of the demonstrative syllogism. The discussion of

definition in effect bore on the middle terms of demonstrative syllogisms. The suggestion is that formally different modes of defining, with respect to removal from matter and motion, ground the formal difference between types of theoretical science. The subject of a demonstration in natural philosophy is defined without singular but with common or universal sensible matter; the subject of a mathematical demonstration is defined without any sensible matter. How can the subject of metaphysics be expressed? The possibility of the science depends on our knowing that some things exist apart from matter and motion. Mathematics does not presuppose the separate existence of its objects; metaphysics does. Why not then say that metaphysics deals with things separated from matter and motion, that is with a particular kind of being? But that is not the subject ever assigned to this effort by Aristotle. The methodological reasons can be found in chapter 17 of Book Seven of the *Metaphysics*: the subject of a science must always be a complex. That is why the subject of this discipline is being as being.

Why should we say that, in our desire to learn more about separate substances, we should take as our subject all the things that are? The short answer is this: in order to be a theology, metaphysics must first be an ontology. Separate substance, divine being, is not directly accessible for our inspection or study. We come upon our first secure knowledge of a god in the proof of the Prime Mover. Tantalizingly, once seen as a necessary requirement for there being any moved movers, the Prime Mover does not become a thematic object of inquiry in natural philosophy. One obvious reason for this is that such an entity is not an instance of the things which fall under the scope of the science. Knowledge of it comes about obliquely and indirectly. The same restriction is operative when the philosopher turns his culminating attention to a deity. How can he know more about the first cause of things? If the Prime Mover is known through moved movers as his effects, any further knowledge of him must be through his effects. It is by describing the effect as widely as possible that one seeks to

come to a knowledge of the first cause unrestricted by the characteristics of mobile things. That characterization is being as being. The subject of metaphysics is being in all its amplitude in order to acquire a knowledge of the cause of being that will be correspondingly unbounded.

11. Philosophical and Scriptural Theology

Earlier we indicated the difference between philosophy and theology in the writings of St. Thomas. That distinction takes theology to mean discourse that takes its rise from the revealed truths of the Bible. But there is also a theology which constitutes the defining telos of philosophical inquiry. In the following passage, Thomas contrasts the two theologies in a way which throws light on what was said in the preceding paragraph.

> Thus it is that divine science or theology is of two kinds, one in which divine things are considered not as the subject of the science but as principles of the subject and this is the theology that the philosophers pursue, also called metaphysics. The other considers divine things in themselves as the subject of the science, and this is the theology which is treated in Sacred Scripture. They are both concerned with things which exist separately from matter and motion, but differently, insofar as they are two ways in which something can exist separately from matter and motion: first, such that it is of the definition of the things said to be separate, that they can never exist in matter in motion, as God and the angels are said to be separate from matter and motion; second, such that it is not part of their definition that they exist in matter and motion, because they can exist apart from matter and motion, although sometimes they are found in matter and motion, for example, substance, potency and act are separate from matter and motion because they do not require matter in order to exist as mathematicals do, although they can be understood without sensible matter.

Philosophical theology treats of things separate in the second way as its subjects and of things separate in the first way as the principles of its subject. But the theology of Sacred Scripture treats of things separate in the first way as its subjects, although in it some things which exist in matter and motion are considered insofar as they are needed to make the divine manifest. (*Exposition of Boethius' On the Trinity*, q. 5, a. 4)

Philosophical theology is not some science distinct from metaphysics; it is simply the name that can be given to metaphysics because it appeals to a god as the cause of its subject. This may make it seem that knowledge of a god is merely a bonus, a tangential consideration; on the contrary, it is the chief aim of the science. But the divine can only be known indirectly, through its effects. For this reason, metaphysics can be viewed as an extended effort to examine substance in order to come to knowledge of the first cause. And given the principle that we name things as we know them (ST Ia.13.1), this can be regarded as a prolonged effort to develop the language with which we speak of a god.

11.1 God

Thomas says that the truth of the proposition *a god exists* is knowable in itself, because the predicate is included in the essence of the subject. But it is not knowable to us, because the essence of a god is unknowable to us. He also says that the essence of a god is existence, that such a being is *ipsum esse subsistens*, and yet that we cannot know this essence. How is any of this coherent? Mustn't one know what one is talking about to deny anything of it, in particular to deny that it is knowable to us? How can Thomas simultaneously assert what the essence of a god is and deny that we know it?

In order to understand why his claims about the existence and essence of a god are not incoherent, we need to place them within the context of Aristotle's *Posterior Analytics*. According to Aristotle, one mode of *per se* predication, the first, is that in which the predicate of the proposition is included within the definition of the subject. We've already seen the second, where the subject is included in the defintion of the predicate, the mode appropriate to the powers of a subject. So in the first mode, if one immediately knew the essential definition of the subject, one would immediately know that a particular proposition is *per se* true simply by knowing that its predicate is included within that essential definition. Any proposition in which the predicate is included in the essential definition of the subject is knowable in itself. For instance, Thomas thinks that anyone who knows the language will know the truth of a proposition like *a whole consists of the sum of its parts*. Because the terms are related in this fashion and so fundamental in the language, no special knowledge is necessary to grasp its truth. Such a proposition is thus knowable in itself but also to us.

However, clearly this account leaves open the possibility of subjects in which the essential definition is either unknown or even unknowable. For instance, if we suppose that H2O is the essential definition of water, we have to recognize that there will be many who will not know it. It will not be immediately "known to us," but require learning. No doubt we can still refer to water in statements about it because the term 'water' has a nominal definition, clear-potable-odorless etc., used by the community to refer to what is in fact H2O. So that *water is H2O* will be knowable in itself, even if unknowable to us, until we engage in Chemistry. Consider the mind. Clearly we use the term 'mind' meaningfully in any number of sentences. But perhaps, as Colin McGinn has argued, the actual nature of mind is incomprehensible to limited minds such as ours. In that case it might be knowable in itself, and yet strictly unknowable to us. Thus the

distinction between what is knowable in itself and what is knowable to us is not incoherent.

What of the claims that the essence of a god is not just unknown to us, but unknowable to us, that the essence of a god is His existence, and that it is *ipsum esse subsistens*? Don't these remain jointly inconsistent and thus incoherent, even if the underlying distinction is not? No. In claiming that the essence of a god is not knowable to us, Thomas is talking about its accessibility to philosophical inquiry. The human mind of itself is proportioned to knowing material things. It can only know immaterial things insofar as causal arguments can be made to posit the existence of such things as necessary to the explanation of material things—causes that are only appealed to when one has excluded the possibility of a material explanation of the phenomenon. But we've already seen that to claim that something is immaterial is not to know any property of it, much less its essence. Still, it remains available to Thomas to claim that while the knowledge of the essence of a god is unknowable to philosophy, it is known to us by Revelation. Indeed, he appeals to God's revelation to Moses on Sinai to establish the claim that God's essence is *ipsum esse subsistens*—that is to say, the being who reveals Himself to Moses identifies Himself as what the philosophers are looking for but cannot find in its essence. And Christians believe that God further discloses Himself in the Incarnation of Christ and the Gospel narratives as a Trinity of Persons in Unity of Substance. (*Summa theologiae* Ia.31.1 ad 4) Here, in knowing the essence of God as trinitarian we have another example, like the Resurrection, of something that can only be known by faith in God's Revelation. It is not something that can be known by both Revelation and Philosophy. So the essence of God is knowable in itself, and also to the learned. But the learned are not the philosophers. Rather they are all those who know it by faith in God's revelation.

So, can the existence of God be philosophically demonstrated? If God's essence is His existence, and His essence remains in principle philosophically unknowable to us, how could it be demonstrated? In fact, Aquinas claims that it can be demonstrated that there is *a* god, and that there is only *one* god. That God's essence remains in principle philosophically unknowable to us is the basis for Aquinas' denial that the existence of God can be demonstrated *a priori*. And any reliance upon knowledge of the essence that is only known to us by faith would by that fact cease to be properly philosophical. However, we have seen that Aquinas relies upon the distinction between nominal definitions of terms and essential definitions of the subjects referred to by those terms. To demonstrate the existence of a god one may use nominal definitions that appeal to a god as the cause of various phenomena. This is to argue *a posteriori*. The appeal to these nominal definitions forms the basis for Aquinas' Five Ways (*Summa Theologiae*, Ia.2.3) all of which end with some claim about how the term 'god' is used.

Again, some will claim that Aquinas isn't really interested in proving the existence of God in these Five Ways. After all, he already knows the existence of God by faith, and he is writing a theological work for beginners. What need is there of proving the existence of something he already knows exists? The Ways are very sketchy, and don't even necessarily conclude to a single being, much less God or the Christian God. In addition, Aquinas claims that God's essence is his existence and that we cannot know His essence, so we cannot know His existence. Aquinas must really intend the Five Ways as less than proofs; they are more like incomplete propaedeutic considerations for thinking adequately about God in Sacred Theology. In effect, Aquinas doesn't think philosophy can in fact demonstrate the existence of God.

But as elsewhere these claims are ambiguous and suffer at the hands of Thomas' own texts. In the first place, the objection that he already knows

by faith that God exists has some merit in it, if we understand it as directed at a reading of Thomas that would have him attempting a foundational enterprise of grounding religious faith in what is rationally demonstrable by Philosophy. But that reading is anachronistic, and does not attend to the context of the *Summa Theologiae*. There is no reason to think that Thomas thinks the proofs are necessary for the rationality of religious faith. They are part of the enterprise of showing that *Sacra Doctrina* meets the condition of a science as described by Aristotle in the *Posterior Analytics*, which he had argued in the first question of the *Summa*, an issue that is different from the question of the broad rationality of religious faith.

In addition, the objections end up denying what Aquinas writes immediately before the Five Ways—that the existence of a god is "demonstrable." (*Summa Theologiae* Ia.2.2) And his introduction of the Five Ways begins by saying that the existence of a god can be "proved" in Five Ways. To counter the objection that he must mean something informal here by "demonstrate" and "prove," one need only recognize the explicit use of Aristotle's *Posterior Analytics* to sort through the question. He cites Aristotle's distinction between demonstrating the existence of some subject, and going on to demonstrate properties of that subject by appeal to the essence of the subject as cause of those properties. The first kind of demonstration is called demonstration *quia*, the second demonstration *propter quid*. In order to have any science at all, the subject matter must exist. So demonstration *quia* must precede demonstration *propter quid*. If you want to have a science of unicorns, you have to show me that there is at least one unicorn to be studied. There is no science of what does not exist. So there are two demonstrative stages in any science, the demonstration of the existence of the subject (*quia*), and the demonstration of the properties of the subject through its essence (*propter quid*). Aquinas' denial that the essence of God can be known philosophically is a denial that one can have *propter quid* scientific understanding of a god through philosophy. It is not a denial that there can

be demonstration *quia* of the existence of a god. There is no reason to deny that Thomas thinks the Five Ways are proofs or demonstrations in the most robust sense, namely that which he appeals to as set out by Aristotle in the *Posterior Analytics*.

Notice however the back and forth between the use of 'God' as a proper name and the use of 'god' as a common noun. One source of the ambiguity in the objections comes about because it is claimed that Aquinas does not think one can demonstrate the existence of God. But in terms of the *Posterior Analytics* one cannot demonstrate the existence of anything under a proper name. One can point at Socrates, and say "see, Socrates is alive." One cannot do that with God. In addition, one cannot give a formal argument for Socrates existence using 'Socrates'. One can only demonstrate in the relevant sense using common nouns, since such nouns are the only ones that have definitions, either nominal or essential. So strictly speaking it is true that Thomas doesn't think one can demonstrate the existence of *God* in the Five Ways. He recognizes the difference between 'God' used as a proper noun, and 'god' used as a common noun. (*Summa Theologiae* Ia.13.9) The ambiguity is pronounced in Latin which lacks the indefinite article 'a', where in English we can disambiguate between 'God' and 'a god'. The situation is exacerbated by translations that simply translate 'deus' in the Ways as 'God' in English. In the Five Ways, he does not use 'god'as a proper name, but as a common noun having five different nominal definitions. So each of the ways concludes that there is "a god." Thus, it is also true that the Five Ways do not as such prove that there is only one god. It is for that reason that Thomas himself thinks one must actually argue additionally that a god must be utterly unique, and thus that there can be only one, which he does several questions after the Five Ways (*Summa Theologiae* Ia.11). Of course, once the utter uniqueness of a god has been shown, one can begin to use "God" as a proper name to refer to that utterly unique being, as he subsequently argues in *Summa Theologiae* Ia.13 does happen.

It is the utter uniqueness and singularity of a god that undermines the objection that whatever the philosophical arguments terminate in, it is not the god of Judaism, Christianity, and Islam, a god who is only known by faith. That is simply to deny Thomas' claim that the god Jews, Christians, and Muslims believe in can be known, but only partially by philosophical analysis. If the demonstrations work, as Thomas thinks they do, what other god would the Jews, Christians, and Muslims believe in? There is no other god. Given God's revelation, He must be identified with the utterly unique being the philosopher's argument terminates in.

Finally, the sketchy character of the Ways reflects the fact that they are directed at beginning students. However the audience of beginners that Thomas has in mind are not beginners in Philosophy. They are beginners in *Sacra Doctrina*. As we have seen, in the medieval educational setting such beginners would be thoroughly steeped in the philosophical disciplines before ever being allowed to study *Sacra Doctrina*. So Aquinas could expect his readers to know the much more extensive and complete arguments he was gesturing at with the Five Ways, arguments to be found in detail in other figures like Aristotle, Avicenna, and so on, as well as in other works of his own, the *Summa Contra Gentiles* for example. In short, even if the Five Ways are judged to be unsound demonstrations, a judgment that requires close analysis and examination of the filled out arguments, there is no reason to suggest that Thomas took them any less seriously as demonstrations or proofs in the fullest sense.

Now, even though there can be no demonstration *propter quid* of God's essence or nature, this does not mean that philosophical theology is left with a bare knowledge of the existence of God, and nothing more. The second stage of science will go on, but it will go on in a mode deeply indebted to Pseudo-Dionysius and Neoplatonism with the approach often called the "via negativa." Instead of arguing positively from the essence of God to His properties, one will argue from God's effects, particularly the

perfections of creatures that do not of necessity involve material embodiment, to the affirmation that God possesses these perfections. However, recognizing that the way in which God possesses these perfections must be different from the way in which creatures possess them, one must deny that God has them in the creaturely mode. Instead He must possess them in a "super eminent" fashion that we cannot comprehend. So, while on the basis of effect to cause arguments we can say that God is just, wise, good, perfect, and so on, we do not know what it is for God to be just, wise, good, and perfect. We end up denying of God the creaturely mode of these perfections. In this way God is approached negatively by denying things of Him rather than by directly knowing what God is. This account relies heavily upon the use of analogous names in talking about God and creatures.

11.2 Analogous Names

Aristotle spoke of "things said in many ways," a notable instance of which is "being." One of the difficulties with assigning being, or being qua being as the subject of a science is that a subject must be univocally common to the things that fall under it. But 'being' is not univocal, as it has a plurality of meanings. Aristotle solved this problem with his account of "things said in many ways," by observing that while they have many meanings, these form an ordered set with one of the meanings as primary and regulative. Substance is being in the primary sense, which is why the science of being as being is effectively a science of substance. Thomas's term for such names is analogy: 'being' is an analogous term and its primary analogate is substance.

In the crucial middle books of the *Metaphysics*—Seven and Eight—we have an analysis of substance which takes off from material substance, which is a compound of matter and form, and arrives at a notion of substance as form alone. This definition does not fit material substance, of

course, but it is devised in order to be able to apply the term substance to the immaterial things whose existence has been established in natural philosophy. This extension of names, whose natural habitat is sensible things, to God is another instance of analogous naming for Thomas. Names common to God and creatures bring out another feature of our knowing. If we ask what the primary analogate of names common to God and creatures is, the answer is: the meaning of the term as it applies to creatures. The word must be refined before it can be applied to God and this means the formation of an extended meaning which leans on the primary meaning for its intelligibility.

Consider the example of 'wise'. Both men and God are said to be wise. What can we mean when we say that God is wise? Not the same thing as when we say that Socrates is wise. Socrates became wise and wisdom is a trait which with age and forgetfulness he could lose. Thus to be Socrates and to be wise are not the same thing. But in the case of God, 'wise' does not signify some incidental property He might or might not have. This is captured by noting that while we say God is wise, we also say he *is* wisdom. This suffices to indicate the way in which the meaning of the term as applied to God involves negating features of its meaning as it applies to men.

If God is thus named secondarily by the common name, so that the creature is primarily named by it, nonetheless God's wisdom is the cause and source of human wisdom. Ontologically, God is primary and the creature secondary. Names analogously common to God and creature thus underscore the way in which what comes to be known first for us is not first in reality, and what is first in reality is not first in our knowledge.

11.3 Essence and Existence

It is evident that material substances exist contingently. They come into being and they pass out of being. While they exist, their existing is not what they are. Thomas accepts from Boethius that it is self-evident that what a thing is and its existing differ (*diversum est esse et id quod est*). Material things depend upon causes to exist, both to become and to be. There is no need to dwell on this except insofar as it provides a springboard to speak of immaterial substance. Only in God is it the case that what he is and his existing are identical: God is His own existence. The phrase Thomas uses to express this is *ipsum esse subsistens*. Of course this is paradoxical. Existence is the actuality of a substance, not itself something subsistent. This is true with material substances. But when we ask what we mean by saying that God exists, we have to negate aspects of material existence in order to avoid speaking of Him as if he were a contingent being.

The problem that Thomas now faces is how to speak of the immaterial substances which are less than God although superior to material substances, that is, angels. For a material thing to exist is for its form actually to inhere in its matter. But what is it for a pure form to exist? Since immaterial substances less than God are dependent on the divine causality in order to exist, existing cannot be what they are of their essence. In short, in angels too there is a distinction of essence and existence. Thomas notes that a created separate substance is what it is and not another thing: that is, it has the perfection it has, but not unlimited perfection. It is a being of a kind, not being as such. Gabriel is perfect as to his nature, but he lacks the perfection of being Raphael or Michael. Form thus operates as a restriction on existence as such. In God alone is there unrestricted existence; He is existence, *ipsum esse subsistens*. Here we have an argument for the fact that God's essence is his existence, because His essence is *not* a restriction of *esse* to a finite expression or character. And yet it remains true that while we know the fact, we do not

know the why of the fact because the knowledge of God's essence remains unknown to us.

12. Moral Doctrine

Thomas' Moral Doctrine is primarily eudaimonistic and virtue based. Human beings always act for an end that is conceived of as good. A desired good provides the motive for initiating and completing some act. Action begins in desire and ends in satisfaction or joy in completion— the achievement or acquisition of the good apprehended and desired. Properly human action proceeds from and is under the control of intellect and will. While human beings have many appetites informed by sense cognition of the world, they also have rational appetite that is informed by an intellectual apprehension of the world and the goods within it appropriate to human flourishing. Errors of apprehension are certainly possible, and yet a human action always originates in the apprehension of some apparent good by intellect and the desire for it by the will informed by the apprehension. Will is rational appetite. Actions are judged to be good or bad in relation to real human goods for which they are either conducive (good) or detrimental (bad). Given the complexity of human life and the goods appropriate to it, it may well happen that a particular action may be judged to be good in many ways, and yet also bad in others. For one to have acted well simply is for one to have done something that is good in every respect. There is one single ultimate human good that provides an ordering of all other human goods as partial in relation to it, namely, happiness or better in the Latin *beatitudo*.

When Aristotle sought to isolate the human good, he employed the so-called function argument. If one knows what a carpenter is or does he has the criteria for recognizing a good carpenter. So too with bank-tellers, golfers, brain surgeons and locksmiths. If then man as such has a function, we will have a basis for deciding whether someone is a good human being.

But what could this function be? Just as we do not appraise carpenters on the basis of their golf game or golfers on the basis of their being able to pick locks, we will not want to appraise the human agent on an incidental basis. So too we do not appraise the carpenter in terms of his weight, the condition of his lungs, or his taste buds. No more would we appraise a human being on the basis of activities similar to those engaged in by non-human animals. The activity that sets the human agent apart from all others is rational activity. The human agent acts knowingly-willingly. If this is the human function, the human being who performs it well will be a good person and be happy.

Thomas argues that there is one single end for all human beings, and that it is happiness. However, that is a formal description of the end, leaving open the material specification of just what that happiness is for a human being. Thomas distinguishes in the *Summa Theologiae* between the imperfect happiness of this life and the perfect happiness of the next life in beatitude or union with God. And on the basis of this distinction some will argue that Aquinas ultimately finds Aristotle's function argument unsatisfying, insofar as the result of the function argument is supposed to be the claim that happiness consists in a complete life lived in accord with reason and virtue. And here again it will be claimed that Aquinas in some sense rejects the fundamentals of the Aristotelian account. Insofar as he describes the life in accord with reason and virtue in this life as imperfect, he must be suggesting that it is in some sense faulty, not true or real happiness. Real happiness is something other.

But such an interpretation fails on a number of counts. In the first place it misunderstands Aquinas' use of 'imperfect' which does not mean *faulty* or *false*. It can mean *not as great by comparison*, as in the claim that human beings are imperfect with regard to the angels. This claim is not meant to suggest that human beings are faulty or false angels; it simply means that their perfection is not as great in the scale of being as that of the angels. It

can also mean incomplete in the constitution of some overall good. So the pursuit of some limited good, say education, is imperfect because not the complete human good, even though it is partially constitutive of the human good. But it is certainly not a faulty or false human good.

In the second place, such a claim about Aquinas has to confront his own understanding of Aristotle. Aquinas claims that Aristotle understood that a complete life in accord with reason and virtue in this life is incomplete or imperfect happiness. (See his commentary on the *Nichomachean Ethics*, Book 1, lect. 16, #200–202). Indeed, Aristotle himself says that perfect happiness is to be associated with the divine. (*Nichomachean Ethics*, 1099b9–13) Thus Aquinas does not claim for himself the distinction between imperfect and perfect happiness, but attributes it to Aristotle. And so his use of it in the *Summa Theologiae* cannot be taken to be a rejection of the analysis Aristotle provides of the formal characteristics of happiness.

Obviously, one may fault Thomas for his understanding of Aristotle. But the claim that he misinterprets Aristotle is no argument that he rejects Aristotle. In fact, his interpretation of Aristotle on imperfect and perfect happiness embodies the thesis he expresses in the *Summa Theologiae* that we saw above. The philosophers are capable of grasping some of the things that are constitutive of or necessary for perfect happiness in beatitude. Revelation concerning even those matters they can grasp is necessary, because what they have grasped takes a long time, is very difficult, and may be filled with errors. God in his mercy makes these things known in revelation in order that perfect happiness may be attained. And yet, Thomas never abandons the fundamental affirmation of the human capacity to understand apart from revelation the nature of happiness in formal terms and what constitutes its imperfect material status in this life, even as its perfect embodiment in the next remains unattainable to philosophy without the resources of faith.

Many have come to this point, pulse quickened by the possibilities of the function-argument, only to be gripped with doubt at this final application of it. Rational activity seems too unmanageable a description to permit a function-analysis of it. Of course Aristotle agrees, having made the point himself. Rational activity is said in many ways or, as Thomas would put it, it is an analogous term. It covers an ordered set of instances. There is the activity of reason as such, there is the activity of reason in its directive or practical capacity, and there are bodily movements and the like which are rational insofar as rational provides the adequate formal description of them. If the virtue of a function is to perform it well, the analogy of "rational activity" makes clear that there is a plurality of virtues. Moral virtues are habits of appetite brought about by the direction of reason. Temperance is to seek pleasure rationally; courage is to react to the threat of harm rationally. The virtues of practical intellect are art and prudence; the virtues of theoretical intellect are insight, science and wisdom.

All this and much more enters into Thomas's moral teaching. Thomas will distinguish acts of a human from human acts, the former being activities truly found in human agents, but also found in other non-human agents too. For example, the act of a human might be as important as the beating of his heart or or as trivial as the nervous tapping of his fingers. The human act is one which proceeds from and is under the control of reason and will. Since the human act by definition is the pursuit of a known good, the question arises as to the relationship between the objects of the myriad acts that humans perform. Is there some over-all good sought by human agents? Is there an ultimate end of human action?

In commenting on chapter two of Book One of the *Nicomachean Ethics* where Aristotle argues for there being an ultimate end, Thomas points out that the argument is actually a series of *reductiones ad absurdum*. That is, the denial of an ultimate end of human action reduces to the claim that there is no end to human seeking at all, that it is pointless. This analysis

has not gotten the attention it deserves: the implication is that it is self-evident that there is an ultimate end which is why denials of it must flounder in incoherence. The argument for an ultimate end that Thomas puts forth in the *Summa Theologiae* is somewhat different. Any action aims at some good. A particular good by definition shares in and is not identical with goodness itself. What binds together all the acts that humans perform is the overarching goodness they seek in this, that, and the other thing. That over arching goodness, what Thomas calls the *ratio bonitatis*, is the ultimate end. It follows that anything a human agent does is done for the sake of the ultimate end.

This dissatisfies because we feel we are owed a richer account of goodness. After all, human agents differ insofar as they have different notions of what goodness is. Fame, wealth, pleasure, power, and so on seem to function as the dominant purpose of different persons. Thomas could scarcely overlook this, let alone deny it. Can his earlier position on the unity of the ultimate end still stand? The fact that there are false or inadequate identifications of goodness does not mean that there is not a true and adequate account of what is perfecting or fulfilling of human agents. Everyone acts on the supposition that what he does will contribute to his overall good; one's overall good is the ultimate reason for doing anything. But not everything one does under this aegis actually contributes to one's overall good. Thus in one sense there is one and the same ultimate end for every human agent—the integral human good—and there are correct and mistaken notions of what actually constitutes this integral good.

This may seem like an empty claim, but it provides a basis on which to proceed. If indeed every human agent acts for the sake of his overall good, the discussion can turn to whether or not what he here and now pursues, or his general theory of what constitutes the overall good, can withstand scrutiny. It is not necessary to persuade anyone that he ought to pursue the

ultimate end in the sense of his overall good. What else would he pursue? But if one is persuaded that what he pursues does not contribute to his overall good, he already has reasons for changing his ways.

12.1 Virtue

Thomas' broad account of virtues as excellences or perfections of the various human powers formally echoes Aristotle, both with regard to the nature of a virtue and many specific virtues. Virtues are developed habits of powers disposing agents to good actions. Because human actions are those acts that are subject to the rule of human reason and will, the human virtues reside in the various powers that are subject to the rule of reason and will. As developed dispositions they stand as intermediate states between the powers simply and the full blown exercises of those powers. A virtue is a kind of mean between excess and defect in the exercise of a power. For example, with regard to eating a temperate person eats what is appropriate to him or her, pursuing neither too much which would lead to gluttony nor to little which would lead to starvation.

However, it is a mistake to think that Thomas simply repeats Aristotle. In the first place, he discusses a wide array of both particular virtues and vices never discussed by Aristotle. More importantly the theological setting of his work results in a number of distinctions concerning the virtues that pertain to human life and happiness. Thomas first distinguishes a twofold happiness for human beings. One is the sort of happiness that is achievable by a human agent in this life through the exercise of the powers he or she is endowed with by nature. He says this is a happiness that is "proportionate" to human nature. It may well include the sort of contemplation of the nature of divinity the philosophers aspire to. However, Thomas adds that there is "another" happiness that cannot be achieved simply by the exercise of the human powers without divine supernatural assistance. (*Summa theologiae* Ia-IIae.62.1) This is a

happiness not to be found perfectly in this life, but only in the next. It is *beatitudo* or blessedness strictly speaking. It consists in the intellectual vision of God and all things "in" God. While the kind of contemplation of God the philosophers aspire to in this life consists in knowing God through His effects and the ways in which they represent Him as cause, this *beatitudo* of the next life Thomas describes as a participation in the life of divinity itself as the essence of God Himself is united by the "light of glory" to the intellect of the human being (*Summa theologiae* Ia.12), a union with God that results in the experience of ultimate and final joy or delight of the will. (*Summa theologiae* IIa-IIae.11.3-4) However, even though this *beatitudo* is brought about supernaturally by the power of God, it is not utterly foreign to human nature. In effect, the supernatural power of God elevates or expands the powers of intellect and will to a kind of completion beyond themselves and yet not foreign to them. So this distinction of a "twofold happiness" should not be thought of as involving two fundamentally distinct goals or ends of human life. The second supernatural happiness is seen as a kind of surpassing perfection of the first.

This distinction of a twofold happiness in human life leads to a distinction between the natural virtues and the theological virtues. Natural virtues are virtues that pertain to the happiness of this life that is "proportionate" to human nature. Theological virtues pertain to the *beatitudo* that is not proportionate to human nature, the supernatural good of life with God. Natural virtues are divided into moral virtues and intellectual virtues. The intellectual virtues perfect the intellect and confer an aptness for the good work of the intellect which is the apprehension of truth. The moral virtues are the habits that perfect the various powers concerned with human appetites, including rational appetite, conferring upon them an aptness for the right use of those appetites.

The cardinal natural virtues are Prudence, Justice, Courage, and Temperance. Prudence is an intellectual virtue since it bears upon the goal of truth in the good ordering of action. In addition, because there are two specific powers of the generic sensitive appetite, the concupiscent and the irascible, there are two cardinal virtues that pertain to them. The concupiscent appetite inclines one toward what is suitable and away from what is harmful to human bodily life. Temperance is the cardinal virtue that pertains to it. The irascible appetite inclines one toward resisting those things that attack human bodily life. Courage is the cardinal virtue that pertains to it. Finally, Justice is a virtue of the rational appetite or will. These virtues are called "cardinal" both because of their specific importance, but also as general headings under which the wide array of particular virtues are classed. Temperance and Courage are ordered toward and perfect the good of the individual as such, while Justice is ordered toward and perfects the good of others in relation to the individual.

The theological virtues are Faith, Hope, and Love. They bear upon eternal beatitude and are simply infused by God's gift of grace. They cannot be acquired by human effort. However, as noted above the "second" supernatural happiness is not foreign to the first natural happiness, but a kind of surpassing perfection of it. So along with the infusion of the theological virtues, Thomas holds that natural virtues are infused along with them. Thus there is a distinction between "infused natural virtues" and "acquired natural virtues." As infused, the natural virtues cannot be acquired by human effort, although they may be strengthened by it. Acquired natural virtues, on the other hand, are the corresponding virtues that can be acquired by human effort without the gift of divine grace. While Thomas acknowledges that these acquired natural virtues can in principle be developed by human effort without grace, he thinks that their actual acquisition by human effort is very difficult due to the influence of sin.

In addition, the infused natural virtues spring from Charity as its effects, and thus bear upon its object, which is the love of God and the love of neighbor in God. A primary example for Thomas is *Misericordia* which is the virtue that pertains to suffering with others and acting to alleviate their suffering. It looks like Justice because it bears upon the good of another. And yet it is different from Justice because it springs from the natural friendship that all human beings bear to one another, and requires that one take upon oneself the sufferings of other human beings. Thomas explicitly but unconvincingly claims that Aristotle recognized it. And yet in the *Summa Theologiae* he says that it is an effect of Charity. In that case there is an acquired form of it and an infused form of it. As infused, it is informed by the love of God and the love of neighbor in God which is beatitude.

The infused natural virtues differ in important respects from the corresponding acquired virtues because as infused they point toward the supernatural end, and the mean in acquired virtue is fixed by human reason while the mean in the infused virtue is according to divine rule. Thomas gives as an example the difference between acquired and infused Temperance. Acquired temperance is a mean inclining a human being to eat enough food to sustain his or her health and not harm the body. Infused temperance is a mean inclining the human being through abstinence to castigate and subject the body.

Even one mortal or grave sin destroys both Charity and all the infused moral virtues that proceed from it, while leaving Hope and Faith as lifeless habits that are no longer virtues. On the other hand, a single sin, whether venial or mortal, does not destroy the acquired natural virtues.

Charity, as we've seen, is the love of God and neighbor in God. It resides in the will. Hope is the desire for the difficult but attainable good of eternal happiness or beatitude. It too resides in the will. Faith is

intellectual assent to revealed supernatural truths that are not evident in themselves or through demonstration from truths evident in themselves. So it resides in the intellect. It is divided into believing that there is a god and other truths pertaining to that truth, believing God, and believing "in" God. The distinction between the last two is subtle. It is one thing to say you believe me. It is a different thing to say you believe in me. The latter connotes the relation of your intellect to the will's desire to direct yourself to me in love. Thus believing in God goes well beyond believing that there is a god. It suggests the other theological virtues of Charity and Hope.

In beatitude and felicity, the fulfillment of intellect and will respectively, the virtues of Faith and Hope fall away, and do not exist, for one now sees with the intellect what one believed, and has attained what one hoped for with the will. Only Charity abides.

12.2 Natural Law

Thomas's reading of Aristotle's argument for the ultimate end as a *reductio* and his own claim that in one sense of it everyone pursues the ultimate end since one chooses whatever he chooses *sub ratione boni* and as conducive to or a constituent of his fulfillment and perfection, tell us something important about Thomas's mode of procedure. We said earlier that philosophy begins from pre-philosophical principles already had by everyone. In the moral order, it is essential that one uncover the starting point, the latent presupposition of any action, clarify it and proceed from there. This procedure is equally manifest in Thomas's treatment of what he calls natural law.

What is natural law? One description of it is: the peculiarly human participation in the eternal law, in providence. All creatures are ordered to an end, have natures whose fulfillment is what it is because of those natures. It is not peculiar to man that he is fashioned so as to find his good

in the fulfillment of his nature. That is true of anything. But other things are ordered to ends of which they themselves are not conscious. It is peculiar to man that he becomes aware of the good and freely directs himself to it. Of course man is not free to choose the good—any choice is a choice under the aspect of the good. And as to what is really as opposed to only apparently his good, he is not free to make that what it is. He is, however, free to direct himself or not to his true end.

A second description of natural law is: the first principles or starting points of practical reasoning. To indicate what he means by this, Thomas invokes the analogy of the starting points of reasoning as such. We have already mentioned the distinction between knowledge of the simple and knowledge of the complex. The former is a concept and is expressed in a definition or description. The latter is an affirmation or negation of one thing of another. There is something which is first in each of these orders. That is, Thomas holds that there is a conception which is prior to and presupposed by all other conceptions and a judgment that is prior to and presupposed by all other judgments. Since knowledge is expressed by language, this seems to come down to the assertion that there is a first word that everyone utters and a first statement that would appear in everyone's baby book on the appropriate page. But surely that is false. So what does Thomas mean?

He says that our first conception is of being, of that which is, and our first judgment is that you cannot affirm and deny the same thing in the same sense simultaneously. Since few if any humans first utter 'being' or its equivalent and no one fashions as his first enunciation the principle of contradiction, facts as known to Thomas as ourselves, his meaning must be more subtle. It is that whatever concept one first forms and expresses verbally—Mama, hot, whatever,is a specification or an instance of that which is, *being*. Aristotle had observed that children at first call all men father and all women mother. The terms then function as generic for any

male or female. Even more basically, each presupposes that what is generically grasped is an instance of being. Being is prior not because it is grasped absolutely, without reference to this being or that. It is some particular being that is first of all grasped, and however it is named it will mean minimally something that is.

So too with regard to the first judgment. Children express their recognition of this principle when they disagree over the location of some quite specific thing, say a baseball mitt. One accuses the other of taking it. You did. I didn't. You did. I didn't. A fundamental disagreement. But what they are agreed on is that if it were true that one did it could not simultaneously and in the same sense be true that he did not. The principle is latent in, implicit in, any concrete judgment just as being is involved in any other conception.

It is on an analogy with these starting points of thinking as such that Thomas develops what he means by natural law. In the practical order there is a first concept analogous to being in the theoretical order and it is the good. The good means what is sought as fulfilling of the seeker. The first practical judgment is: the good should be done and pursued and evil avoided. Any other practical judgment is a specification of this one and thus includes it. Natural Law consists of this first judgment and other most general ones that are beyond contest. These will be fashioned with reference to constituents of our complete good—existence, food, drink, sex and family, society, desire to know. We have natural inclinations to such goods. Natural law precepts concerning them refer the objects of natural inclinations to our overall or integral good, which they specify.

Most moral judgments are true, if true, only by and large. They express means to achieve our overall good. But because there is not a necessary connection between the means and end, they can hold only for the most part. Thus there are innumerable ways in which human beings lead their

lives in keeping with the ultimate end. Not all means are necessarily related to the end. Moral philosophy reposes on natural law precepts as common presuppositions, but its advice will be true only in the main. So the lives of human beings will show a great deal of variation in the ways they pursue the human end in accord with these general principles. Thus the need for the virtues bearing upon the contingencies of life, Prudence in particular.

It might be noted that when Thomas, following Aristotle, says that man is by nature a social or political animal, he does not mean that each of us has a tendency to enter into social contracts or the like. The natural in this sense is what is not chosen, but given, and what is given about human life is that we are in the first place born into the community of the family, are dependent on it for years in order to survive, and that we flourish as human beings within various larger social and political communities. The moral consists in behaving well in these given settings.

13. Thomism

Thomas's teaching came under attack, largely by Franciscans, immediately after his death. Dominicans responded. This had the effect of making Dominicans Thomists and Franciscans non–Thomists—Bonaventurians, Scotists, Ockhamists. The Jesuits were founded after the Reformation and they tended to be Thomists, often with a Suarezian twist.

When in 1879 Leo XIII issued the encyclical *Aeterni Patris* calling for the revival of the study of Thomas Aquinas, he was not directing his readers to one school as opposed to others. Thomas was put forward as the paladin of philosophy in its true sense, as over and against the vagaries of modern thought since Descartes. The response to Leo's call was global and sustained. New journals and learned societies were founded, curricula were reshaped to benefit from the thought of Thomas and this not simply

in seminaries and pontifical universities but throughout the world in colleges and universities. Such giants as Jacques Maritain and Etienne Gilson may be taken to symbolize the best of this Thomistic revival.

Vatican II, the ecumenical council that met from 1962–1965 drew this stage of the Thomistic Revival to a close. It was widely held that the Council had dethroned Thomas in favor of unnamed contemporary philosophers. (When they were named, quarrels began.) In the post-conciliar period, Catholics have adopted many contemporary philosophical trends with mixed results, as the speed with which such trends come and go has appeared to accelerate, without obvious lasting results. Now with the vogue of the notion that modernity has failed and the Enlightenment Project come a cropper, many, Catholics and non-Catholics alike, are turning to Thomas as a spur or foil for their thinking. In 1998 John Paul II issued an encyclical called *Fides et Ratio*. In its reaffirmation of the importance of Thomas, it may be regarded as the charter of the Thomism of the third millennium.

Bibliography

Primary Literature

Martin, Christopher, 1988. *The Philosophy of Thomas Aquinas: Introductory Readings*. London: Routledge Kegan & Paul.
McDermott, Timothy, 1993. *Aquinas Selected Writings*. New York: Oxford University Press.
McInerny, Ralph, 1998. *Thomas Aquinas Selected Writings*. London: Penguin Classics.

Life and Works

Weisheipl, James A., 1974. *Thomas D'Aquino: His Life, Thought and*

Enlightenment

First published Fri Aug 20, 2010; substantive revision Tue Aug 29, 2017

The heart of the eighteenth century Enlightenment is the loosely organized activity of prominent French thinkers of the mid-decades of the eighteenth century, the so-called "*philosophes*"(e.g., Voltaire, D'Alembert, Diderot, Montesquieu). The *philosophes* constituted an informal society of men of letters who collaborated on a loosely defined project of Enlightenment exemplified by the project of the *Encyclopedia* (see below 1.5). However, there are noteworthy centers of Enlightenment outside of France as well. There is a renowned Scottish Enlightenment (key figures are Frances Hutcheson, Adam Smith, David Hume, Thomas Reid), a German Enlightenment (*die Aufklärung*, key figures of which include Christian Wolff, Moses Mendelssohn, G.E. Lessing and Immanuel Kant), and there are also other hubs of Enlightenment and Enlightenment thinkers scattered throughout Europe and America in the eighteenth century.

What makes for the unity of such tremendously diverse thinkers under the label of "Enlightenment"? For the purposes of this entry, the Enlightenment is conceived broadly. D'Alembert, a leading figure of the French Enlightenment, characterizes his eighteenth century, in the midst of it, as "the century of philosophy *par excellence*", because of the tremendous intellectual and scientific progress of the age, but also because of the expectation of the age that philosophy (in the broad sense of the time, which includes the natural and social sciences) would dramatically improve human life. Guided by D'Alembert's characterization of his century, the Enlightenment is conceived here as having its primary origin in the scientific revolution of the 16th and 17th centuries. The rise of the new science progressively undermines not only the ancient geocentric conception of the cosmos, but also the set of presuppositions that had served to constrain and guide philosophical inquiry in the earlier times.

The dramatic success of the new science in explaining the natural world promotes philosophy from a handmaiden of theology, constrained by its purposes and methods, to an independent force with the power and authority to challenge the old and construct the new, in the realms both of theory and practice, on the basis of its own principles. Taking as the core of the Enlightenment the aspiration for intellectual progress, and the belief in the power of such progress to improve human society and individual lives, this entry includes descriptions of relevant aspects of the thought of earlier thinkers, such as Hobbes, Locke, Descartes, Bayle, Leibniz, and Spinoza, thinkers whose contributions are indispensable to understanding the eighteenth century as "the century of philosophy *par excellence*".

The Enlightenment is often associated with its political revolutions and ideals, especially the French Revolution of 1789. The energy created and expressed by the intellectual foment of Enlightenment thinkers contributes to the growing wave of social unrest in France in the eighteenth century. The social unrest comes to a head in the violent political upheaval which sweeps away the traditionally and hierarchically structured *ancien régime* (the monarchy, the privileges of the nobility, the political power of the Catholic Church). The French revolutionaries meant to establish in place of the *ancien régime* a new reason-based order instituting the Enlightenment ideals of liberty and equality. Though the Enlightenment, as a diverse intellectual and social movement, has no definite end, the devolution of the French Revolution into the Terror in the 1790s, corresponding, as it roughly does, with the end of the eighteenth century and the rise of opposed movements, such as Romanticism, can serve as a convenient marker of the end of the Enlightenment, conceived as an historical period.

For Enlightenment thinkers themselves, however, the Enlightenment is not an historical period, but a process of social, psychological or spiritual development, unbound to time or place. Immanuel Kant defines

"enlightenment" in his famous contribution to debate on the question in an essay entitled "An Answer to the Question: What is Enlightenment?" (1784), as humankind's release from its self-incurred immaturity; "immaturity is the inability to use one's own understanding without the guidance of another." Expressing convictions shared among Enlightenment thinkers of widely divergent doctrines, Kant identifies enlightenment with the process of undertaking to think for oneself, to employ and rely on one's own intellectual capacities in determining what to believe and how to act. Enlightenment philosophers from across the geographical and temporal spectrum tend to have a great deal of confidence in humanity's intellectual powers, both to achieve systematic knowledge of nature and to serve as an authoritative guide in practical life. This confidence is generally paired with suspicion or hostility toward other forms or carriers of authority (such as tradition, superstition, prejudice, myth and miracles), insofar as these are seen to compete with the authority of one's own reason and experience. Enlightenment philosophy tends to stand in tension with established religion, insofar as the release from self-incurred immaturity in this age, daring to think for oneself, awakening one's intellectual powers, generally requires opposing the role of established religion in directing thought and action. The faith of the Enlightenment – if one may call it that – is that the process of enlightenment, of becoming progressively self-directed in thought and action through the awakening of one's intellectual powers, leads ultimately to a better, more fulfilled human existence.

This entry describes the main tendencies of Enlightenment thought in the following main sections: (1) The True: Science, Epistemology, and Metaphysics in the Enlightenment; (2) The Good: Political Theory, Ethical Theory and Religion in the Enlightenment; (3) The Beautiful: Aesthetics in the Enlightenment.

- 1. The True: Science, Epistemology and Metaphysics in the

1. The True: Science, Epistemology and Metaphysics in the Enlightenment

In this era dedicated to human progress, the advancement of the natural sciences is regarded as the main exemplification of, and fuel for, such progress. Isaac Newton's epochal accomplishment in his *Principia Mathematica* (1687), which, very briefly described, consists in the comprehension of a diversity of physical phenomena – in particular the motions of heavenly bodies, together with the motions of sublunary bodies – in few relatively simple, universally applicable, mathematical laws, was

a great stimulus to the intellectual activity of the eighteenth century and served as a model and inspiration for the researches of a number of Enlightenment thinkers. Newton's system strongly encourages the Enlightenment conception of nature as an orderly domain governed by strict mathematical-dynamical laws and the conception *of ourselves* as capable of knowing those laws and of plumbing the secrets of nature through the exercise of our unaided faculties. – The conception of nature, and of how we know it, changes significantly with the rise of modern science. It belongs centrally to the agenda of Enlightenment philosophy to contribute to the new knowledge of nature, and to provide a metaphysical framework within which to place and interpret this new knowledge.

1.1 Rationalism and the Enlightenment

René Descartes' rationalist system of philosophy is one of the pillars on which Enlightenment thought rests. Descartes (1596–1650) undertakes to establish the sciences upon a secure metaphysical foundation. The famous method of doubt Descartes employs for this purpose exemplifies (in part through exaggerating) an attitude characteristic of the Enlightenment. According to Descartes, the investigator in foundational philosophical research ought to doubt all propositions that can be doubted. The investigator determines whether a proposition is dubitable by attempting to construct a possible scenario under which it is false. In the domain of fundamental scientific (philosophical) research, no other authority but one's own conviction is to be trusted, and not one's own conviction either, until it is subjected to rigorous skeptical questioning. With his method, Descartes casts doubt upon the senses as authoritative source of knowledge. He finds that God and the immaterial soul are both better known, on the basis of innate ideas, than objects of the senses. Through his famous doctrine of the dualism of mind and body, that mind and body are two distinct substances, each with its own essence, the material world (allegedly) known through the senses becomes denominated as an

"external" world, insofar as it is external to the ideas with which one immediately communes in one's consciousness. Descartes' investigation thus establishes one of the central epistemological problems, not only of the Enlightenment, but also of modernity: the problem of objectivity in our empirical knowledge. If our evidence for the truth of propositions about extra-mental material reality is always restricted to mental content, content before the mind, how can we ever be certain that the extra-mental reality is not other than we represent it as being? Descartes' solution depends on our having secured prior and certain knowledge of God. In fact, Descartes argues that *all* human knowledge (not only knowledge of the material world through the senses) depends on metaphysical knowledge of God.

Despite Descartes' grounding of all scientific knowledge in metaphysical knowledge of God, his system contributes significantly to the advance of natural science in the period. He attacks the long-standing assumptions of the scholastic-aristotelians whose intellectual dominance stood in the way of the development of the new science; he developed a conception of matter that enabled mechanical explanation of physical phenomena; and he developed some of the fundamental mathematical resources – in particular, a way to employ algebraic equations to solve geometrical problems – that enabled the physical domain to be explained with precise, simple mathematical formulae. Furthermore, his grounding of physics, and all knowledge, in a relatively simple and elegant rationalist metaphysics provides a model of a rigorous and complete secular system of knowledge. Though major Enlightenment thinkers (for example Voltaire in his *Letters on the English Nation*, 1734) embrace Newton's physical system in preference to Descartes', Newton's system itself depends on Descartes' earlier work, a dependence to which Newton himself attests.

Cartesian philosophy also ignites various controversies in the latter decades of the seventeenth century that provide the context of intellectual tumult out of which the Enlightenment springs. Among these

controversies are the following: Are mind and body really two distinct sorts of substances, and if so, what is the nature of each, and how are they related to each other, both in the human being (which presumably "has" both a mind and a body) and in a unified world system? If matter is inert (as Descartes claims), what can be the source of motion and the nature of causality in the physical world? And of course the various epistemological problems: the problem of objectivity, the role of God in securing our knowledge, the doctrine of innate ideas, and others.

Baruch Spinoza's systematic rationalist metaphysics, which he develops in his *Ethics* (1677) in part in response to problems in the Cartesian system, is also an important basis for Enlightenment thought. Spinoza develops, in contrast to Cartesian dualism, an ontological monism according to which there is only one substance, God or nature, with two attributes, corresponding to mind and body. Spinoza's denial, on the basis of strict philosophical reasoning, of the existence of a transcendent supreme being, his identification of God with nature, gives strong impetus to the strands of atheism and naturalism that thread through Enlightenment philosophy. Spinoza's rationalist principles also lead him to assert a strict determinism and to deny any role to final causes or teleology in explanation. (See Israel 2001.)

The rationalist metaphysics of Leibniz (1646–1716) is also foundational for the Enlightenment, particularly the German Enlightenment (*die Aufklärung*), one prominent expression of which is the Leibnizian rationalist system of Christian Wolff (1679–1754). Leibniz articulates, and places at the head of metaphysics, the great rationalist principle, the principle of sufficient reason, which states that everything that exists has a sufficient reason for its existence. This principle exemplifies the characteristic conviction of the Enlightenment that the universe is thoroughly rationally intelligible. The question arises of how this principle itself can be known or grounded. Wolff attempts to derive it from the

logical principle of non-contradiction (in his *First Philosophy or Ontology*, 1730). Criticism of this alleged derivation gives rise to the general question of how formal principles of logic can possibly serve to ground substantive knowledge of reality. Whereas Leibniz exerts his influence through scattered writings on various topics, some of which elaborate plans for a systematic metaphysics which are never executed by Leibniz himself, Wolff exerts his influence on the German Enlightenment through his development of a rationalist system of knowledge in which he attempts to demonstrate all the propositions of science from first principles, known a priori.

Wolff's rationalist metaphysics is characteristic of the Enlightenment by virtue of the *pretensions* of human reason within it, not by reason's success in establishing its claims. Much the same could be said of the great rationalist philosophers of the seventeenth century. Through their articulation of the *ideal* of scientia, of a complete science of reality, composed of propositions derived demonstratively from a priori first principles, these philosophers exert great influence on the Enlightenment. But they fail, rather spectacularly, to realize this ideal. To the contrary, what they bequeath to the eighteenth century is metaphysics, in the words of Kant, as "a battlefield of endless controversies." However, the controversies themselves – regarding the nature of God, mind, matter, substance, cause, et cetera, and the relations of each of these to the others – provide tremendous fuel to Enlightenment thought.

1.2 Empiricism and the Enlightenment

Despite the confidence in and enthusiasm for human reason in the Enlightenment – it is sometimes called "the Age of Reason" – the rise of empiricism, both in the practice of science and in the theory of knowledge, is characteristic of the period. The enthusiasm for reason in the Enlightenment is primarily not for the faculty of reason as an independent

source of knowledge, which is embattled in the period, but rather for the human cognitive faculties generally; the Age of Reason contrasts with an age of religious faith, not with an age of sense experience. Though the great seventeenth century rationalist metaphysical systems of Descartes, Spinoza and Leibniz exert tremendous influence on philosophy in the Enlightenment; moreover, and though the eighteenth-century Enlightenment has a rationalist strain (perhaps best exemplified by the system of Christian Wolff), nevertheless, that the *Encyclopedia* of Diderot and D'Alembert is dedicated to three empiricists (Francis Bacon, John Locke and Isaac Newton), signals the ascendency of empiricism in the period.

If the founder of the rationalist strain of the Enlightenment is Descartes, then the founder of the empiricist strain is Francis Bacon (1561–1626). Though Bacon's work belongs to the Renaissance, the revolution he undertook to effect in the sciences inspires and influences Enlightenment thinkers. The Enlightenment, as the age in which experimental natural science matures and comes into its own, admires Bacon as "the father of experimental philosophy." Bacon's revolution (enacted in, among other works, *The New Organon*, 1620) involves conceiving the new science as (1) founded on empirical observation and experimentation; (2) arrived at through the method of induction; and (3) as ultimately aiming at, and as confirmed by, enhanced practical capacities (hence the Baconian motto, "knowledge is power").

Of these elements of Bacon's revolution, the point about method deserves special emphasis. Isaac Newton's work, which stands as the great exemplar of the accomplishments of natural science for the eighteenth century, is, like Bacon's, based on the inductive method. Whereas rationalist of the seventeenth century tend to conceive of scientific knowledge of nature as consisting in a system in which statements expressing the observable phenomena of nature are *deduced* from first

principles, known a priori, Newton's method begins with the observed phenomena of nature and reduces its multiplicity to unity by induction, that is, by finding mathematical laws or principles from which the observed phenomena can be derived or explained. The evident success of Newton's "bottom-up" procedure contrasts sharply with the seemingly endless and fruitless conflicts among philosophers regarding the meaning and validity of first principles of reason, and this contrast naturally favors the rise of the Newtonian (or Baconian) method of acquiring knowledge of nature in the eighteenth century.

The tendency of natural science toward progressive independence from metaphysics in the eighteenth century is correlated with this point about method. The rise of modern science in the sixteenth and seventeenth centuries proceeds through its separation from the presuppositions, doctrines and methodology of theology; natural science in the eighteenth century proceeds to separate itself from metaphysics as well. Newton proves the capacity of natural science to succeed independently of a priori, clear and certain first principles. The characteristic Enlightenment suspicion of all allegedly authoritative claims the validity of which is obscure, which is directed first of all against religious dogmas, extends to the claims of metaphysics as well. While there are significant Enlightenment thinkers who are metaphysicians – again, one thinks of Christian Wolff – the general thrust of Enlightenment thought is anti-metaphysical.

John Locke's *Essay Concerning Human Understanding* (1690) is another foundational text of the Enlightenment. A main source of its influence is the epistemological rigor that it displays, which is at least implicitly anti-metaphysical. Locke undertakes in this work to examine the human understanding in order to determine the limits of human knowledge; he thereby institutes a prominent pattern of Enlightenment epistemology. Locke finds the source of all our ideas, the ideas out of which human

knowledge is constructed, in the senses and argues influentially against the rationalists' doctrine of innate ideas. Locke's sensationalism exerts great influence in the French Enlightenment, primarily through being taken up and radicalized by the *philosophe*, Abbé de Condillac. In the *Treatise on Sensations* (1754), Condillac attempts to explain how all human knowledge arises out of sense experience. Locke's epistemology, as developed by Condillac and others, contributes greatly to the emerging science of psychology in the period.

Locke and Descartes both pursue a method in epistemology that brings with it the epistemological problem of objectivity. Both examine our knowledge by way of examining the ideas we encounter directly in our consciousness. This method comes to be called "the way of ideas". Though neither for Locke nor for Descartes do *all* of our ideas represent their objects by way of *resembling* them (e.g., our idea of God does not represent God by virtue of resembling God), our alleged knowledge of our environment through the senses does depend largely on ideas that allegedly resemble external material objects. The way of ideas implies the epistemological problem of how we can know that these ideas do in fact resemble their objects. How can we be sure that these objects do not appear one way before the mind and exist in another way (or not at all) in reality outside the mind? George Berkeley, an empiricist philosopher influenced by John Locke, avoids the problem by asserting the metaphysics of idealism: the (apparently material) objects of perception are nothing but ideas before the mind. However, Berkeley's idealism is less influential in, and characteristic of, the Enlightenment, than the opposing positions of materialism and Cartesian dualism. Thomas Reid, a prominent member of the Scottish Enlightenment, attacks the way of ideas and argues that the immediate objects of our (sense) perception are the common (material) objects in our environment, not ideas in our mind. Reid mounts his defense of naïve realism as a defense of common sense over against the doctrines of the philosophers. The defense of common

sense, and the related idea that the results of philosophy ought to be of use to common people, are characteristic ideas of the Enlightenment, particularly pronounced in the Scottish Enlightenment.

1.3 Skepticism in the Enlightenment

Skepticism enjoys a remarkably strong place in Enlightenment philosophy, given that confidence in our intellectual capacities to achieve systematic knowledge of nature is a leading characteristic of the age. This oddity is at least softened by the point that much skepticism in the Enlightenment is merely methodological, a tool meant to serve science, rather than a position embraced on its own account. The instrumental role for skepticism is exemplified prominently in Descartes' *Meditations on First Philosophy* (1641), in which Descartes employs radical skeptical doubt to attack prejudices derived from learning and from sense experience and to search out principles known with certainty which may serve as a secure foundation for a new system of knowledge. Given the negative, critical, suspicious attitude of the Enlightenment towards doctrines traditionally regarded as well founded, it is not surprising that Enlightenment thinkers employ skeptical tropes (drawn from the ancient skeptical tradition) to attack traditional dogmas in science, metaphysics and religion.

However, skepticism is not merely a methodological tool in the hands of Enlightenment thinkers. The skeptical cast of mind is one prominent manifestation of the Enlightenment spirit. The influence of Pierre Bayle, another founding figure of the Enlightenment, testifies to this. Bayle was a French Protestant, who, like many European philosophers of his time, was forced to live and work in politically liberal and tolerant Holland in order to avoid censorship and prison. Bayle's *Historical and Critical Dictionary* (1697), a strange and wonderful book, exerts great influence on the age. The form of the book is intimidating: a biographical dictionary, with long scholarly entries on obscure figures in the history of culture, interrupted by

long scholarly footnotes, which are in turn interrupted by further footnotes. Rarely has a work with such intimidating scholarly pretentions exerted such radical and liberating influence in the culture. It exerts this influence through its skeptical questioning of religious, metaphysical, and scientific dogmas. Bayle's eclecticism and his tendency to follow arguments without pre-arranging their conclusions make it difficult to categorize his thought. It is the attitude of inquiry that Bayle displays, rather than any doctrine he espouses, that mark his as distinctively Enlightenment thought. He is fearless and presumptuous in questioning all manner of dogma. His attitude of inquiry resembles both that of Descartes' meditator and that of the person undergoing enlightenment as Kant defines it, the attitude of coming to think for oneself, of daring to know. This epistemological attitude, as manifest in distrust of authority and reliance on one's own capacity to judge, expresses the Enlightenment values of individualism and self-determination.

This skeptical/critical attitude underlies a significant tension in the age. While it is common to conceive of the Enlightenment as supplanting the authority of tradition and religious dogma with the authority of reason, in fact the Enlightenment is characterized by a crisis of authority regarding any belief. This is perhaps best illustrated with reference to David Hume's skepticism, as developed in Book One of *A Treatise of Human Nature* (1739–40) and in his later *Enquiries Concerning Human Understanding* (1748). While one might take Hume's skepticism to imply that he is an outlier with respect to the Enlightenment, it is more convincing to see Hume's skepticism as a flowering of a crisis regarding authority in belief that is internal to the Enlightenment. Hume articulates a variety of skepticisms. His "skepticism with regard to the senses" is structured by the epistemological problem bound up with the way of ideas, described above. Hume also articulates skepticism with regard to reason in an argument that is anticipated by Bayle. Hume begins this argument by noting that, though rules or principles in demonstrative sciences are certain or infallible, given

the fallibility of our faculties, our applications of such rules or principles in demonstrative inferences yield conclusions that cannot be regarded as certain or infallible. On reflection, our conviction in the conclusions of demonstrative reasoning must be qualified by an assessment of the likelihood that we made a mistake in our reasoning. Thus, Hume writes, "all knowledge degenerates into probability" (*Treatise*, I.iv.i). Hume argues further that, given this degeneration, for any judgment, our assessment of the likelihood that we made a mistake, and the corresponding diminution of certainty in the conclusion, is another judgment about which we ought make a further assessment, which leads to a further diminution of certainty in our original conclusion, leading "at last [to] a total extinction of belief and evidence". Hume also famously questions the justification of inductive reasoning and causal reasoning. According to Hume's argument, since in causal reasoning we take our past observations to serve as evidence for judgments regarding what will happen in relevantly similar circumstances in the future, causal reasoning depends on the assumption that the future course of nature will resemble the past; and there is no non-circular justification of this essential assumption. Hume concludes that we have no rational justification for our causal or inductive judgments. Hume's skeptical arguments regarding causal reasoning are more radical than his skeptical questioning of reason as such, insofar as they call into question even *experience itself* as a ground for knowledge and implicitly challenge the credentials of Newtonian science itself, the very pride of the Enlightenment. The question implicitly raised by Hume's powerful skeptical arguments is whether *any* epistemological authority at all can withstand critical scrutiny. The Enlightenment begins by unleashing skepticism in attacking limited, circumscribed targets, but once the skeptical genie is out of the bottle, it becomes difficult to maintain conviction in any authority. Thus, the despairing attitude that Hume famously expresses in the conclusion to Book One of the *Treatise*, as the consequence of his epistemological

inquiry, while it clashes with the self-confident and optimistic attitude we associate with the Enlightenment, in fact reflects an essential possibility in a distinctive Enlightenment problematic regarding authority in belief.

1.4 Science of Man and Subjectivism in the Enlightenment

Though Hume finds himself struggling with skepticism in the conclusion of Book One of the *Treatise*, the project of the work as he outlines it is not to advance a skeptical viewpoint, but to establish a science of the mind. Hume is one of many Enlightenment thinkers who aspire to be the "Newton of the mind"; he aspires to establish the basic laws that govern the elements of the human mind in its operations. Alexander Pope's famous couplet in *An Essay on Man* (1733) ("Know then thyself, presume not God to scan/ The proper study of mankind is man") expresses well the intense interest humanity gains in itself within the context of the Enlightenment, as a partial substitute for its traditional interest in God and the transcendent domain. Just as the sun replaces the earth as the center of our cosmos in Copernicus' cosmological system, so humanity itself replaces God at the center of humanity's consciousness in the Enlightenment. Given the Enlightenment's passion for science, the self-directed attention naturally takes the form of the rise of the scientific study of humanity in the period.

The enthusiasm for the scientific study of humanity in the period incorporates a tension or paradox concerning the place of humanity in the cosmos, as the cosmos is re-conceived in the context of Enlightenment philosophy and science. Newton's success early in the Enlightenment of subsuming the phenomena of nature under universal laws of motion, expressed in simple mathematical formulae, encourages the conception of nature as a very complicated machine, whose parts are material and whose motions and properties are fully accounted for by deterministic causal laws. But if our conception of nature is of an exclusively material domain

governed by deterministic, mechanical laws, and if we at the same time deny the place of the supernatural in the cosmos, then how does humanity itself fit into the cosmos? On the one hand, the achievements of the natural sciences in general are the great pride of the Enlightenment, manifesting the excellence of distinctively human capacities. The pride and self-assertiveness of humanity in the Enlightenment expresses itself, among other ways, in humanity's making the study of itself its central concern. On the other hand, the study of humanity in the Enlightenment typically yields a portrait of us that is the opposite of flattering or elevating. Instead of being represented as occupying a privileged place in nature, as made in the image of God, humanity is represented typically in the Enlightenment as a fully natural creature, devoid of free will, of an immortal soul, and of a non-natural faculty of intelligence or reason. The very title of J.O. de La Mettrie's *Man a Machine* (1748), for example, seems designed to deflate humanity's self-conception, and in this respect it is characteristic of the Enlightenment "science of man". It is true of a number of works of the Enlightenment, perhaps especially works in the more radical French Enlightenment – notable here are Helvétius's *Of the Spirit* (1758) and Baron d'Holbach's *System of Nature* (1770) – that they at once express the remarkable self-assertiveness of humanity characteristic of the Enlightenment in their scientific aspirations while at the same time painting a portrait of humanity that dramatically deflates its traditional self-image as occupying a privileged position in nature.

The methodology of epistemology in the period reflects a similar tension. Given the epistemological role of Descartes' famous "*cogito, ergo sum*" in his system of knowledge, one might see Descartes' epistemology as already marking the transition from an epistemology privileging knowledge of God to one that privileges self-knowledge instead. However, in Descartes' epistemology, it remains true that knowledge of God serves as the necessary foundation for all human knowledge. Hume's *Treatise* displays such a re-orientation less ambiguously. As noted, Hume means

his work to comprise a science of the mind or of man. In the Introduction, Hume describes the science of man as effectively a foundation for all the sciences since all sciences "lie under the cognizance of men, and are judged of by their powers and faculties." In other words, since all science is human knowledge, scientific knowledge *of humanity* is the foundation of the sciences. Hume's placing the science of man at the foundation of all the sciences both exemplifies the privilege afforded to "mankind's study of man" within the Enlightenment and provides an interpretation of it. But Hume's methodological privileging of humanity in the system of sciences contrasts sharply with what he says in the body of his science about humanity. In Hume's science of man, reason as a faculty of knowledge is skeptically attacked and marginalized; reason is attributed to other animals as well; belief is shown to be grounded in custom and habit; and free will is denied. So, even as knowledge of humanity supplants knowledge of God as the keystone of the system of knowledge, the scientific perspective on humanity starkly challenges humankind's self-conception as occupying a privileged position in the order of nature.

Immanuel Kant explicitly enacts a revolution in epistemology modeled on the Copernican in astronomy. As characteristic of Enlightenment epistemology, Kant, in his *Critique of Pure Reason* (1781, second edition 1787) undertakes both to determine the limits of our knowledge, and at the same time to provide a foundation of scientific knowledge of nature, and he attempts to do this by examining our human faculties of knowledge critically. Even as he draws strict limits to rational knowledge, he attempts to defend reason as a faculty of knowledge, as playing a necessary role in natural science, in the face of skeptical challenges that reason faces in the period. According to Kant, scientific knowledge of nature is not merely knowledge of what *in fact* happens in nature, but knowledge of the causal laws of nature according to which what in fact happens *must* happen. But how is knowledge of necessary causal connection in nature possible? Hume's investigation of the idea of cause had made clear that we cannot

know causal necessity through experience; experience teaches us at most what in fact happens, not what *must* happen. In addition, Kant's own earlier critique of principles of rationalism had convinced him that the principles of ("general") logic also cannot justify knowledge of *real* necessary connections (in nature); the formal principle of non-contradiction can ground at best the deduction of one *proposition* from another, but not the claim that one *property* or *event* must follow from another in the course of nature. The generalized epistemological problem Kant addresses in the *Critique of Pure Reason* is: how is science possible (including natural science, mathematics, metaphysics), given that all such knowledge must be (or include) knowledge of real, substantive (not merely logical or formal) necessities. Put in the terms Kant defines, the problem is: how is synthetic, a priori knowledge possible?

According to Kant's Copernican Revolution in epistemology addressed to this problem, objects must conform themselves to human knowledge rather than knowledge to objects. Certain cognitive forms lie ready in the human mind – prominent examples are the pure concepts of substance and cause and the forms of intuition, space and time; given sensible representations must conform themselves to these forms in order for human experience (as empirical knowledge of nature) to be possible at all. We can acquire scientific knowledge of nature because we constitute it a priori according to certain cognitive forms; for example, we can know nature as a causally ordered domain because we originally synthesize a priori the given manifold of sensibility according to the category of causality, which has its source in the human mind.

Kant saves rational knowledge of nature by limiting rational knowledge to nature. According to Kant's argument, we can have rational knowledge only of the domain of possible experience, not of supersensible objects such as God and the soul. Moreover Kant's solution brings with it a kind of idealism: given the mind's role in constituting objects of experience, we

know objects only as *appearances*, only as they appear according to our faculties, not as they are in themselves. This is the subjectivism of Kant's epistemology. Kant's epistemology exemplifies Enlightenment thought by replacing the theocentric conception of knowledge of the rationalist tradition with an anthropocentric conception.

However, Kant means his system to make room for humanity's practical and religious aspirations toward the transcendent as well. According to Kant's idealism, the realm of nature is limited to a realm of appearances, and we can intelligibly think supersensible objects such as God, freedom and the soul, though we cannot know them. Through the postulation of a realm of unknowable noumena (things in themselves) over against the realm of nature as a realm of appearances, Kant manages to make place for practical concepts that are central to our understanding of ourselves even while grounding our scientific knowledge of nature as a domain governed by deterministic causal laws. Though Kant's idealism is highly controversial from its initial publication, a main point in its favor, according to Kant himself, is that it reconciles, in a single coherent tension, the main tension between the Enlightenment's conception of nature, as ordered according to deterministic causal laws, and the Enlightenment's conception of ourselves, as morally free, as having dignity, and as perfectible.

1.5 Emerging Sciences and the *Encyclopedia*

The commitment to careful observation and description of phenomena as the starting point of science, and then the success at explaining and accounting for observed phenomena through the method of induction, naturally leads to the development of new sciences for new domains in the Enlightenment. Many of the human and social sciences have their origins in the eighteenth century (e.g., history, anthropology, aesthetics, psychology, economics, even sociology), though most are only formally

established as autonomous disciplines later. The emergence of new sciences is aided by the development of new scientific tools, such as models for probabilistic reasoning, a kind of reasoning that gains new respect and application in the period. Despite the multiplication of sciences in the period, the ideal remains to comprehend the diversity of our scientific knowledge as a unified system of science; however, this ideal of unity is generally taken as regulative, as an ideal to emerge in the ever-receding end-state of science, rather than as enforced from the beginning by regimenting science under a priori principles.

As exemplifying these and other tendencies of the Enlightenment, one work deserves special mention: the *Encyclopedia*, edited by Denis Diderot and Jean La Rond d'Alembert. The *Encyclopedia* (subtitled: "*systematic dictionary of the sciences, arts and crafts*") was published in 28 volumes (17 of text, 11 of plates) over 21 years (1751–1772), and consists of over 70,000 articles, contributed by over 140 contributors, among them many of the luminaries of the French Enlightenment. The work aims to provide a compendium of existing human knowledge to be transmitted to subsequent generations, a transmission intended to contribute to the progress and dissemination of human knowledge and to a positive transformation of human society. The orientation of the *Encyclopedia* is decidedly secular and implicitly anti-authoritarian. Accordingly, the French state of the *ancien régime* censors the project, and it is completed only through the persistence of Diderot. The collaborative nature of the project, especially in the context of state opposition, contributes significantly to the formation of a shared sense of purpose among the wide variety of intellectuals who belong to the French Enlightenment. The knowledge contained in the *Encyclopedia* is self-consciously social both in its production – insofar as it is immediately the product of what the title page calls "a society of men of letters" – and in its address – insofar as it is primarily meant as an instrument for the education and improvement of society. It is a striking feature of the *Encyclopedia*, and one by virtue of

which it exemplifies the Baconian conception of science characteristic of the period, that its entries cover the whole range and scope of knowledge, from the most abstract theoretical to the most practical, mechanical and technical.

2. The Good: Political Theory, Ethical Theory and Religion in the Enlightenment

2.1 Political Theory

The Enlightenment is most identified with its political accomplishments. The era is marked by three political revolutions, which together lay the basis for modern, republican, constitutional democracies: The English Revolution (1688), the American Revolution (1775–83), and the French Revolution (1789–99). The success at explaining and understanding the natural world encourages the Enlightenment project of re-making the social/political world, in accord with the models we allegedly find in our reason. Enlightenment philosophers find that the existing social and political orders do not withstand critical scrutiny. Existing political and social authority is shrouded in religious myth and mystery and founded on obscure traditions. The criticism of existing institutions is supplemented with the positive work of constructing in theory the model of institutions as they ought to be. We owe to this period the basic model of government founded upon the consent of the governed; the articulation of the political ideals of freedom and equality and the theory of their institutional realization; the articulation of a list of basic individual human rights to be respected and realized by any legitimate political system; the articulation and promotion of toleration of religious diversity as a virtue to be respected in a well ordered society; the conception of the basic political powers as organized in a system of checks and balances; and other now-familiar features of western democracies. However, for all the enduring accomplishments of Enlightenment political philosophy, it is not clear that

human reason proves powerful enough to put a concrete, positive authoritative ideal in place of the objects of its criticism. As in the epistemological domain, reason shows its power more convincingly in criticizing authorities than in establishing them. Here too the question of the limits of reason is one of the main philosophical legacies of the period. These limits are arguably vividly illustrated by the course of the French Revolution. The explicit ideals of the French Revolution are the Enlightenment ideals of individual freedom and equality; but, as the revolutionaries attempt to devise rational, secular institutions to put in place of those they have violently overthrown, eventually they have recourse to violence and terror in order to control and govern the people. The devolution of the French Revolution into the Reign of Terror is perceived by many as proving the emptiness and hypocrisy of Enlightenment reason, and is one of the main factors which account for the end of the Enlightenment as an historical period.

The political revolutions of the Enlightenment, especially the French and the American, were informed and guided to a significant extent by prior political philosophy in the period. Though Thomas Hobbes, in his *Leviathan* (1651), defends the absolute power of the political sovereign, and is to that extent opposed to the revolutionaries and reformers in England, this work is a founding work of Enlightenment political theory. Hobbes' work originates the modern social contract theory, which incorporates Enlightenment conceptions of the relation of the individual to the state. According to the general social contract model, political authority is grounded in an agreement (often understood as ideal, rather than real) among individuals, each of whom aims in this agreement to advance his rational self-interest by establishing a common political authority over all. Thus, according to the general contract model (though this is more clear in later contract theorists such as Locke and Rousseau than in Hobbes himself), political authority is grounded not in conquest, natural or divinely instituted hierarchy, or in obscure myths and traditions,

but rather in the rational consent of the governed. In initiating this model, Hobbes takes a naturalistic, scientific approach to the question of how political society ought to be organized (against the background of a clear-eyed, unsentimental conception of human nature), and thus decisively influences the Enlightenment process of secularization and rationalization in political and social philosophy.

Baruch Spinoza also greatly contributes to the development of Enlightenment political philosophy in its early years. The metaphysical doctrines of the *Ethics* (1677) lay the groundwork for his influence on the age. Spinoza's arguments against Cartesian dualism and in favor of substance monism, the claim in particular that there can only be one substance, God or nature, was taken to have radical implications in the domains of politics, ethics and religion throughout the period. Spinoza's employment of philosophical reason leads to the denial of the existence of a transcendent, creator, providential, law-giving God; this establishes the opposition between the teachings of philosophy, on the one hand, and the traditional orienting practical beliefs (moral, religious, political) of the people, on the other hand, an opposition that is one important aspect of the culture of the Enlightenment. In his main political work, *Tractatus Theologico-Politicus* (1677), Spinoza, building on his rationalist naturalism, opposes superstition, argues for toleration and the subordination of religion to the state, and pronounces in favor of qualified democracy. Liberalism is perhaps the most characteristic political philosophy of the Enlightenment, and Spinoza, in this text primarily, is one of its originators.

However, John Locke's *Second Treatise of Government* (1690) is the classical source of modern liberal political theory. In his *First Treatise of Government*, Locke attacks Robert Filmer's *Patriarcha* (1680), which epitomizes the sort of political theory the Enlightenment opposes. Filmer defends the right of kings to exercise absolute authority over their subjects

on the basis of the claim that they inherit the authority God vested in Adam at creation. Though Locke's assertion of the natural freedom and equality of human beings in the *Second Treatise* is starkly and explicitly opposed to Filmer's view, it is striking that the cosmology underlying Locke's assertions is closer to Filmer's than to Spinoza's. According to Locke, in order to understand the nature and source of legitimate political authority, we have to understand our relations in the state of nature. Drawing upon the natural law tradition, Locke argues that it is evident to our natural reason that we are all absolutely subject to our Lord and Creator, but that, in relation to each other, we exist naturally in a state of equality "wherein all the power and jurisdiction is reciprocal, no one having more than another" (*Second Treatise*, §4). We also exist naturally in a condition of freedom, insofar as we may do with ourselves and our possessions as we please, within the constraints of the fundamental law of nature. The law of nature "teaches all mankind … that, being all equal and independent, no one ought to harm another in his life, health, liberty, or possessions" (§6). That we are governed in our natural condition by such a substantive moral law, legislated by God and known to us through our natural reason, implies that the state of nature is not Hobbes' war of all against all. However, since there is lacking any human authority over all to judge of disputes and enforce the law, it is a condition marred by "inconveniencies", in which possession of natural freedom, equality and possessions is insecure. According to Locke, we rationally quit this natural condition by contracting together to set over ourselves a political authority, charged with promulgating and enforcing a single, clear set of laws, for the sake of guaranteeing our natural rights, liberties and possessions. The civil, political law, founded ultimately upon the consent of the governed, does not cancel the natural law, according to Locke, but merely serves to draw that law closer. "[T]he law of nature stands as an eternal rule to all men" (§135). Consequently, when established political power violates that law, the people are justified in overthrowing it. Locke's

argument for the right to revolt against a government that opposes the purposes for which legitimate government is taken by some to justify the political revolution in the context of which he writes (the English revolution) and, almost a hundred years later, by others to justify the American revolution as well.

Though Locke's liberalism has been tremendously influential, his political theory is founded on doctrines of natural law and religion that are not nearly as evident as Locke assumes. Locke's reliance on the natural law tradition is typical of Enlightenment political and moral theory. According to the natural law tradition, as the Enlightenment makes use of it, we can know through the use of our unaided reason that we all – all human beings, universally – stand in particular moral relations to each other. The claim that we can apprehend through our unaided reason a *universal* moral order exactly because moral qualities and relations (in particular human freedom and equality) belong to the nature of things, is attractive in the Enlightenment for obvious reasons. However, as noted above, the scientific apprehension of nature in the period does not support, and in fact opposes, the claim that the alleged moral qualities and relations (or, indeed, that *any* moral qualities and relations) are *natural*. According to a common Enlightenment assumption, as humankind clarifies the laws of nature through the advance of natural science and philosophy, the true moral and political order will be revealed with it. This view is expressed explicitly by the *philosophe* Marquis de Condorcet, in his *Sketch for a Historical Picture of the Progress of the Human Mind* (published posthumously in 1795 and which, perhaps better than any other work, lays out the paradigmatically Enlightenment view of history of the human race as a continual progress to perfection). But, in fact, advance in knowledge of the laws of nature in the science of the period does not help with discernment of a natural political or moral order. This asserted relationship between natural scientific knowledge and the political and moral order is under great stress already in the Enlightenment. With respect to Lockean

liberalism, though his assertion of the moral and political claims (natural freedom, equality, et cetera) continues to have considerable force for us, the grounding of these claims in a religious cosmology does not. The question of how to ground our claims to natural freedom and equality is one of the main philosophical legacies of the Enlightenment.

The rise and development of liberalism in Enlightenment political thought has many relations with the rise of the mercantile class (the bourgeoisie) and the development of what comes to be called "civil society", the society characterized by work and trade in pursuit of private property. Locke's *Second Treatise* contributes greatly to the project of articulating a political philosophy to serve the interests and values of this ascending class. Locke claims that the end or purpose of political society is the preservation and protection of property (though he defines property broadly to include not only external property but life and liberties as well). According to Locke's famous account, persons acquire rightful ownership in external things that are originally given to us all by God as a common inheritance, independently of the state and prior to its involvement, insofar as we "mix our labor with them". The civil freedom that Locke defines, as something protected by the force of political laws, comes increasingly to be interpreted as the freedom to trade, to exchange without the interference of governmental regulation. Within the context of the Enlightenment, economic freedom is a salient interpretation of the individual freedom highly valued in the period. Adam Smith, a prominent member of the Scottish Enlightenment, describes in his *An Inquiry into the Nature and Causes of the Wealth of Nations* (1776) some of the laws of civil society, as a sphere distinct from political society as such, and thus contributes significantly to the founding of political economy (later called merely "economics"). His is one of many voices in the Enlightenment advocating for free trade and for minimal government regulation of markets. The trading house floor, in which people of various nationalities, languages, cultures, religions come together and trade, each in pursuit of

his own self-interest, but, through this pursuit, supplying the wants of their respective nations and increasing its wealth, represents for some Enlightenment thinkers the benign, peaceful, universal rational order that they wish to see replace the violent, confessional strife that characterized the then-recent past of Europe.

However, the liberal conception of the government as properly protecting economic freedom of citizens and private property comes into conflict in the Enlightenment with the value of democracy. James Madison confronts this tension in the context of arguing for the adoption of the U.S. Constitution (in his Federalist #10). Madison argues that popular government (pure democracy) is subject to the evil of factions; in a pure democracy, a majority bound together by a private interest, relative to the whole, has the capacity to impose its particular will on the whole. The example most on Madison's mind is that those without property (the many) may seek to bring about governmental re-distribution of the property of the propertied class (the few), perhaps in the name of that other Enlightenment ideal, equality. If, as in Locke's theory, the government's protection of an individual's freedom is encompassed within the general end of protecting a person's property, then, as Madison argues, the proper form of the government cannot be pure democracy, and the will of the people must be officially determined in some other way than by directly polling the people.

Jean-Jacques Rousseau's political theory, as presented in his *On the Social Contract* (1762), presents a contrast to the Lockean liberal model. Though commitment to the political ideals of freedom and equality constitutes a common ground for Enlightenment political philosophy, it is not clear not only how these values have a home in nature as Enlightenment science re-conceives it, but also how concretely to interpret each of these ideals and how properly to balance them against each other. Contrary to Madison, Rousseau argues that direct (pure) democracy is the only form of

government in which human freedom can be realized. Human freedom, according to Rousseau's interpretation, is possible only through governance according to what he calls "the general will," which is the will of the body politic, formed through the original contract, concretely determined in an assembly in which all citizens participate. Rousseau's account intends to avert the evils of factions by structural elements of the original contract. The contract consists in the self-alienation by each associate of all rights and possessions to the body politic. Because each alienates all, each is an equal member of the body politic, and the terms and conditions are the same for all. The emergence of factions is avoided insofar as the good of each citizen is, and is understood to be, equally (because wholly) dependent on the general will. Legislation supports this identification with the general will by preserving the original equality established in the contract, prominently through maintaining a measure of economic equality. Rousseau's account of the ideal relation of the individual citizen to the state differs from Locke's; in Rousseau's account, the individual must be actively engaged in political life in order to maintain the identification of his supremely authoritative will with the general will, whereas in Locke the emphasis is on the limits of governmental authority with respect to the expressions of the individual will. Though Locke's liberal model is more representative of the Enlightenment in general, Rousseau's political theory, which in some respects presents a revived classical model modified within the context of Enlightenment values, in effect poses many of the enduring questions regarding the meaning and interpretation of political freedom and equality within the modern state.

Both Madison and Rousseau, like most political thinkers of the period, are influenced by Baron de Montesquieu's *The Spirit of the Laws* (1748), which is one of the founding texts of modern political theory. Though Montesquieu's treatise belongs to the tradition of liberalism in political theory, given his scientific approach to social, legal and political systems,

his influence extends beyond this tradition. Montesquieu argues that the system of legislation for a people varies appropriately with the particular circumstances of the people. He provides specific analysis of how climate, fertility of the soil, population size, et cetera, affect legislation. He famously distinguishes three main forms of governments: republics (which can either be democratic or aristocratic), monarchies and despotisms. He describes leading characteristics of each. His argument that functional democracies require the population to possess civic virtue in high measure, a virtue that consists in valuing public good above private interest, influences later Enlightenment theorists, including both Rousseau and Madison. He describes the threat of factions to which Madison and Rousseau respond in different (indeed opposite) ways. He provides the basic structure and justification for the balance of political powers that Madison later incorporates into the U.S. Constitution.

It is striking how unenlightened many of the Enlightenment's celebrated thinkers are concerning issues of race and of gender (regarding race, see *Race and Enlightenment: A Reader*, edited by Emmanuel Chukwudi Eze). For all the public concern with the allegedly universal "rights of man" in the Enlightenment, the rights of women and of non-white people are generally overlooked in the period. (Mary Wollstonecraft's *Vindication of the Rights of Woman* (1792) is a noteworthy exception.) When Enlightenment thinkers do turn their attention to the social standing of women or of non-white people, they tend to spout unreasoned prejudice. Moreover, while the philosophies of the Enlightenment generally aspire or pretend to universal truth, unattached to particular time, place or culture, Enlightenment writings are rife with rank ethno- and Eurocentrism, often explicit.

In the face of such tensions within the Enlightenment, one response is to affirm the power of the Enlightenment to improve humanity and society long beyond the end of the eighteenth century, indeed, down to the present

day and into the future. This response embraces the Enlightenment and interprets more recent emancipation movements and achievement of recognition of the rights and dignity of traditionally oppressed and marginalized groups as expressions of Enlightenment ideals and aspirations. Critics of the Enlightenment respond differently to such tensions. Critics see them as symptoms of disorder, ideology, perversity, futility or falsehood that afflict the very core of the Enlightenment itself. (See James Schmidt's "What Enlightenment Project?" for discussion of critics of the Enlightenment.) Famously, Adorno and Horkheimer interpret Nazi death camps as the result of "the dialectic of the Enlightenment", as what historically becomes of the supremacy of instrumental reason asserted in the Enlightenment. As another example, we may point to some post-modern feminists, who argue, in opposition to the liberal feminists who embrace broadly Enlightenment ideals and conceptions, that the essentialism and universalism associated with Enlightenment ideals are both false and intrinsically hostile to the aspirations to self-realization of women and of other traditionally oppressed groups. (See Strickland and the essays in Akkerman and Stuurman.) This entry is not the place to delineate strains of opposition to the Enlightenment, but it is worth noting that post-Enlightenment social and political struggles to achieve equality or recognition for traditionally marginalized or oppressed groups are sometimes self-consciously grounded in the Enlightenment and sometimes marked by explicit opposition to the Enlightenment's conceptions or presuppositions.

2.2 Ethical Theory

Many of the leading issues and positions of contemporary philosophical ethics take shape within the Enlightenment. Prior to the Enlightenment in the West, ethical reflection begins from and orients itself around religious doctrines concerning God and the afterlife. The highest good of humanity, and, accordingly, the content and grounding of moral duties, are conceived

in immediately religious terms. During the Enlightenment, this changes, certainly within philosophy, but to some significant degree, within the population of western society at large. As the processes of industrialization, urbanization, and dissemination of education advance in this period, happiness in this life, rather than union with God in the next, becomes the highest end for more and more people. Also, the violent religious wars that bloody Europe in the early modern period motivate the development of secular, this-worldly ethics, insofar as they indicate the failure of religious doctrines concerning God and the afterlife to establish a stable foundation for ethics. In the Enlightenment, philosophical thinkers confront the problem of developing ethical systems on a secular, broadly naturalistic basis for the first time since the rise of Christianity eclipsed the great classical ethical systems. However, the changes in our understanding of nature and cosmology, effected by modern natural science, make recourse to the systems of Plato and Aristotle problematic. The Platonic identification of the good with the real and the Aristotelian teleological understanding of natural things are both difficult to square with the Enlightenment conception of nature. The general philosophical problem emerges in the Enlightenment of how to understand the source and grounding of ethical duties, and how to conceive the highest good for human beings, within a secular, broadly naturalistic context, and within the context of a transformed understanding of the natural world.

In ethical thought, as in political theory, Hobbes' thought is an important provocation in the Enlightenment. Hobbes understands what is good, as the end of human action, to be "whatsoever is the object of any man's appetite or desire," and evil to be "the object of his hate, and aversion," "there being nothing simply and absolutely so; nor any common rule of good and evil, to be taken from the nature of the objects themselves" (*Leviathan*, chapter 6). Hobbes' conception of human beings as fundamentally motivated by their perception of what is in their own best interest implies the challenge, important for Enlightenment moral

philosophy, to construct moral duties of justice and benevolence out of such limited materials. The basis of human action that Hobbes posits is immediately intelligible and even shared with other animals to some extent; a set of moral duties constructed on this basis would also be intelligible, de-mystified, and fit within the larger scheme of nature. Bernard Mandeville is sometimes grouped with Hobbes in the Enlightenment, especially by critics of them both, because he too, in his popular *Fable of the Bees; or, Private Vices, Public Benefits* (1714), sees people as fundamentally motivated by their perceived self-interest, and then undertakes to tell a story about how moral virtue, which involves conquering one's own appetite and serving the interests of others, can be understood to arise on this basis.

Samuel Clarke, an influential rationalist British thinker early in the Enlightenment, undertakes to show in his *Discourse concerning the Unchangeable Obligations of Natural Religion* (1706), against Hobbes, that the absolute difference between moral good and moral evil lies in the immediately discernible nature of things, independently of any compacts or positive legislation by God or human beings. Clarke writes that "in men's dealing … one with another, it is undeniably more fit, absolutely and in the nature of the thing itself, that all men should endeavor to promote the universal good and welfare of all; than that all men should be continually contriving the ruin and destruction of all". Likewise for the rest of what morality enjoins upon us. According to Clarke, that some actions (those we call morally good or required) are "fit to be done" and others not fit is grounded upon the immediately evident relations in which things stand to each other in nature, just as "the proportions of lines or numbers" are evident to the rational perception of a reasonable being. Similarly, Christian Wolff's rationalist practical philosophy also grounds moral duties in an objective rational order. However, the objective quality on which moral requirements are grounded for Wolff is not the "fitness" of things to be done but rather their perfection. Wolff counts as a founder of

the *Aufklärung* in part because of his attempted derivation of ethical duties from an order of perfection in things, discernable through reason, independently of divine commands.

Rationalist ethics so conceived faces the following obstacles in the Enlightenment. First, as implied above, it becomes increasingly implausible that the objective, mind-independent order is really as rationalist ethicists claim it to be. Second, even if the objective realm were ordered as the rationalist claims, it remains unclear how this order gives rise (on its own, as it were) to obligations binding on our wills. David Hume famously exposes the fallacy of deriving a prescriptive statement (that one ought to perform some action) from a description of how things stand in relation to each other in nature. Prima facie, there is a gap between the rationalist's objective order and a set of prescriptions binding on our wills; if a supreme legislator must be re-introduced in order to make the conformity of our actions to that objective order binding on our wills, then the alleged existence of the objective moral order does not do the work the account asks of it in the first place.

Alongside the rationalist strand of ethical philosophy in the Enlightenment, there is also a very significant empiricist strand. Empirical accounts of moral virtue in the period are distinguished, both by grounding moral virtue on an empirical study of human nature, and by grounding cognition of moral duties and moral motivation in human sensibility, rather than in reason. The Third Earl of Shaftesbury, author of the influential work *Characteristics of Men, Manners, Opinions, Times* (1711), is a founding figure of the empiricist strand. Shaftesbury, like Clarke, is provoked by Hobbes' egoism to provide a non-egoistic account of moral virtue. Shaftesbury conceives the core notion of the goodness of things teleologically: something is good if it contributes to the well-being or furtherance of the system of which it is a part. Individual animals are members of species, and therefore they are good as such insofar as they

contribute to the well-being of the species of which they are a part. Thus, the good of things, including human beings, for Shaftesbury as for Clarke, is an objective quality that is knowable through reason. However, though we can know what is good through reason, Shaftesbury maintains that reason alone is not sufficient to motivate human action. Shaftesbury articulates the structure of a distinctively human moral sensibility. Moral sensibility depends on the faculty of reflection. When we reflect on first-order passions such as gratitude, kindness and pity, we find ourselves approving or liking them and disapproving or disliking their opposites. By virtue of our receptivity to such feelings, we are capable of virtue and have a sense of right and wrong. In this way, Shaftesbury defines the moral sense that plays a significant role in the theories of subsequent Enlightenment thinkers such as Francis Hutcheson and David Hume.

In the rationalist tradition, the conflict within the breast of the person between the requirements of morality and self-interest is canonically a conflict between the person's reason and her passions. Shaftesbury's identification of a moral sentiment in the nature of humanity renders this a conflict within sensibility itself, a conflict between different sentiments, between a self-interested sentiment and an unegoistic sentiment. Though both Shaftesbury and Hutcheson, no less than Clarke, oppose Hobbes's egoism, it is nonetheless true that the doctrine of moral sensibility softens moral demands, so to speak. Doing what is morally right or morally good is intrinsically bound up with a distinctive kind of pleasure on their accounts. It is significant that both Shaftesbury and Hutcheson, the two founders of modern moral sense theory, articulate their ethical theory in conjunction with an aesthetic theory. Arguably the pleasure we feel in the apprehension of something beautiful is *disinterested pleasure*. Our susceptibility to aesthetic pleasure can be taken to reveal that we apprehend and respond to objective (or, anyway, universal) values, not only or necessarily on the basis of reason, but through our natural sensibility instead. Thus, aesthetics, as Shaftesbury and Hutcheson

independently develop an account of it, gives encouragement to their doctrines of moral sensibility. But an account of moral virtue, unlike aesthetics, requires an account of moral *motivation*. As noted above, both Shaftesbury and Hutcheson want to do justice to the idea that proper moral motivation is not the pursuit of pleasure, even disinterested pleasure, but rather an immediate response to the perception of moral value. The problem of giving a satisfying account of moral motivation is a difficult one for empiricist moral philosophers in the Enlightenment.

While for Shaftesbury, at the beginning of the moral sense tradition, moral sense tracks a mind-independent order of value, David Hume, motivated in part by a more radical empiricism, is happy to let the objective order go. We have no access through reason to an independent order of value which moral sense would track. For Hume, morality is founded completely on our sentiments. Hume is often regarded as the main originator of so-called "ethical subjectivism", according to which moral judgments or evaluations (regarding actions or character) do not make claims about independent facts but merely express the subject's feelings or attitudes with respect to actions or character. Such subjectivism is relieved of the difficult task of explaining how the objective order of values belongs to the natural world as it is being reconceived by natural science in the period; however, it faces the challenge of explaining how error and disagreement in moral judgments and evaluations are possible. Hume's account of the standards of moral judgment follows that of Hutcheson in relying centrally on the "natural" responses of an ideal observer or spectator.

Hume's ethics is exemplary of philosophical ethics in the Enlightenment by virtue of its belonging to the attempt to provide a new, empirically grounded science of human nature, free of theological presuppositions. As noted above, the attempts by the members of the French Enlightenment to present a new understanding of human nature are strongly influenced by Locke's "sensationalism", which, radicalized by Condillac, amounts to the

attempt to base all contents and faculties of the human mind on the senses. Typically, the French *philosophes* draw more radical or iconoclastic implications from the new "science of man" than English or Scottish Enlightenment figures. Claude-Adrien Helvétius (1715–1771) is typical here. In *De l'ésprit* (1758), Helvétius follows the Lockean sensationalism of Condillac and pairs it with the claim that human beings are motivated in their actions only by the natural desire to maximize their own pleasure and minimize their pain. *De l'ésprit*, though widely read, gives rise to strong negative reactions in the time, both by political and religious authorities (the Sorbonne, the Pope and the Parlement of Paris all condemn the book) and by prominent fellow *philosophes*, in great part because Helvétius's psychology seems to critics to render moral imperatives and values without basis, despite his best attempts to derive them. Helvétius attempts to ground the moral equality of all human beings by portraying all human beings, whatever their standing in the social hierarchy, whatever their special talents and gifts, as equally products of the nature we share plus the variable influences of education and social environment. But, to critics, Helvétius's account portrays all human beings as equal only by virtue of portraying all as equally worthless (insofar as the claim to equality is grounded on all being equally determined by external factors). However, Helvétius's ideas, in *De l'ésprit* as well as in its posthumously published sequel *De l'homme* (1772), exert a great deal of influence, especially his case for the role of pleasure and pain in human motivation and the role of education and social incentives in shaping individuals into contributors to the social good. Helvétius is sometimes regarded as the father of modern utilitarianism through his articulation of the greatest happiness principle and through his influence on Bentham.

Helvétius is typical in the respect that he is radical in the revisions he proposes, not in common moral judgments or customs of the time, but rather regarding the philosophical grounding of those judgments and customs. But there are some philosophers in the Enlightenment who are

radical in the revisions they propose regarding the content of ethical judgments themselves. The Marquis de Sade is merely the most notorious example, among a set of Enlightenment figures (including also the Marquis de Argens and Diderot himself in some of his writings) who, within the context of the new naturalism and its emphasis on the pursuit of pleasure, celebrate the avid pursuit of sexual pleasure and explicitly challenge the sexual mores, as well as the wider morality, of their time. The more or less fictionalized, philosophically self-conscious "libertine" is one significant expression of Enlightenment ethical thought.

If the French Enlightenment tends to advance this-worldly happiness as the highest good for human beings more insistently than the Enlightenment elsewhere, then Rousseau's voice is, in this as in other respects, a discordant voice in that context. Rousseau advances the cultivation and realization of *human freedom* as the highest end for human beings and thereby gives expression to another side of Enlightenment ethics. As Rousseau describes it, the capacity for individual self-determination puts us in a problematic relation to our natural desires and inclinations and to the realm of nature generally, insofar as that realm is constituted by mechanistic causation. Though Rousseau places a great deal of emphasis on human freedom, and makes significant contributions to our understanding of ourselves as free, he does not address very seriously the problem of the place of human freedom in the cosmos as it is conceived within the context of Enlightenment naturalism.

However, Rousseau's writings help Kant to the articulation of a practical philosophy that addresses many of the tensions in the Enlightenment. Kant follows Rousseau, and disagrees with empiricism in ethics in the period, in emphasizing human freedom, rather than human happiness, as the central orienting concept of practical philosophy. Though Kant presents the moral principle as a principle of practical reason, his ethics also disagrees significantly with rationalist ethics in the period. According to Kant,

rationalists such as Wolff, insofar as they take moral prescriptions to follow from an end given to the will(in Wolff's case, the end of perfection), do not understand us as autonomous in our moral activity. Through interpreting the faculty of the will itself as practical reason, Kant understands the moral principle as internally legislated, thus as not only compatible with freedom, but as equivalent to the principle of a free will, as a principle of autonomy. As noted above, rationalists in ethics in the period are challenged to explain how the objective moral order which reason in us allegedly discerns gives rise to valid prescriptions binding on our wills (the gap between *is* and *ought*). For Kant, the moral order is not independent of our will, but rather represents the formal constraints of willing as such. Kant's account thus both avoids the is-ought gap and interprets moral willing as expressive of our freedom.

Moreover, by virtue of his interpretation of the moral principle as the principle of pure practical reason, Kant is able to redeem the ordinary sense of moral requirements as over-riding, as potentially opposed to the claims of one's happiness, and thus as different in kind from the deliverances of prudential reasoning. This ordinary sense of moral requirements is not easily accommodated within the context of Enlightenment empiricism and naturalism. Kant's stark dichotomy between a person's practical reason and her sensible nature is strongly criticized, both by the subsequent Romantic generation and in the contemporary context; but this dichotomy is bound up with an important benefit of Kant's view – much promoted by Kant himself – within the context of the Enlightenment. Elaborated in the context of Kant's idealism as a contrast between the "realm of freedom" and the "realm of nature", the dichotomy enables Kant's proposed solution to the conflict between freedom and nature that besets Enlightenment thought. As noted above, Kant argues that the application of the causal principle is restricted to the realm of nature, thus making room for freedom, compatibly with the causal determination of natural events required by scientific knowledge.

Additionally, Kant attempts to show that morality "leads ineluctably to" religious belief (in the supersensible objects of God and of the immortal soul) while being essentially not founded on religious belief, thus again vindicating the ordinary understanding of morality while still furthering Enlightenment values and commitments.

2.3 Religion and the Enlightenment

Though the Enlightenment is sometimes represented as the enemy of religion, it is more accurate to see it as critically directed against various (arguably contingent) features of religion, such as superstition, enthusiasm, fanaticism and supernaturalism. Indeed the effort to discern and advocate for a religion purified of such features – a "rational" or "natural" religion – is more typical of the Enlightenment than opposition to religion as such. Even Voltaire, who is perhaps the most persistent, powerful, vocal Enlightenment critic of religion, directs his polemic mostly against the Catholic Church in France – "*l'infâme*" in his famous sign-off in his letters, "*Écrasez l'infâme*" ("Crush the infamous") refers to the Church, not to religion as such. However, controversy regarding the truth-value or reasonableness of religious belief in general, Christian belief in particular, and controversy regarding the proper place of religion in society, occupies a particularly central place in the Enlightenment. It's as if the terrible, violent confessional strife in the early modern period in Europe, the bloody drawn-out wars between the Christian sects, was removed to the intellectual arena in the Enlightenment and became a set of more general philosophical controversies.

Alongside the rise of the new science, the rise of Protestantism in western Christianity also plays an important role in generating the Enlightenment. The original Protestants assert a sort of individual liberty with respect to questions of faith against the paternalistic authority of the Church. The "liberty of conscience", so important to Enlightenment thinkers in general,

and asserted against all manner of paternalistic authorities (including Protestant), descends from this Protestant assertion. The original Protestant assertion initiates a crisis of authority regarding religious belief, a crisis of authority that, expanded and generalized and even, to some extent, secularized, becomes a central characteristic of the Enlightenment spirit. The original Protestant assertion against the Catholic Church bases itself upon the authority of scripture. However, in the Enlightenment, the authority of scripture is strongly challenged, especially when taken literally. Developing natural science renders acceptance of a literal version of the Bible increasingly untenable. But authors such as Spinoza (in his *Tractatus Theologico-Politicus*) present ways of interpreting scripture according to its spirit, rather than its letter, in order to preserve its authority and truth, thus contributing to the Enlightenment controversy of whether some rationally purified version of the religion handed down in the culture belongs to the true philosophical representation of the world or not; and, if so, what its content is.

It is convenient to discuss religion in the Enlightenment by presenting four characteristic forms of Enlightenment religion in turn: deism, religion of the heart, fideism and atheism.

Deism. Deism is the form of religion most associated with the Enlightenment. According to deism, we can know by the natural light of reason that the universe is created and governed by a supreme intelligence; however, although this supreme being has a plan for creation from the beginning, the being does not interfere with creation; the deist typically rejects miracles and reliance on special revelation as a source of religious doctrine and belief, in favor of the natural light of reason. Thus, a deist typically rejects the divinity of Christ, as repugnant to reason; the deist typically demotes the figure of Jesus from agent of miraculous redemption to extraordinary moral teacher. Deism is the form of religion fitted to the new discoveries in natural science, according to which the cosmos

displays an intricate machine-like order; the deists suppose that the supposition of God is necessary as the source or author of this order. Though not a deist himself, Isaac Newton provides fuel for deism with his argument in his *Opticks* (1704) that we must infer from the order and beauty in the world to the existence of an intelligent supreme being as the cause of this order and beauty. Samuel Clarke, perhaps the most important proponent and popularizer of Newtonian philosophy in the early eighteenth century, supplies some of the more developed arguments for the position that the correct exercise of unaided human reason leads inevitably to the well-grounded belief in God. He argues that the Newtonian physical system implies the existence of a transcendent cause, the creator God. In his first set of Boyle lectures, *A Demonstration of the Being and Attributes of God* (1705), Clarke presents the metaphysical or "argument *a priori*" for God's existence. This argument concludes from the rationalist principle that whatever exists must have a sufficient reason or cause of its existence to the existence of a transcendent, necessary being who stands as the cause of the chain of natural causes and effects. Clarke also supports the empirical argument from design, the argument that concludes from the evidence of order in nature to the existence of an intelligent author of that order. In his second set of Boyle lectures, *A Discourse Concerning the Unchangeable Obligations of Natural Religion* (1706), Clarke argues as well that the moral order revealed to us by our natural reason requires the existence of a divine legislator and an afterlife, in which the supreme being rewards virtue and punishes vice. In his Boyle lectures, Clarke argues directly against the deist philosophy and maintains that what he regards as the one true religion, Christianity, is known as such on the basis of miracles and special revelation; still, Clarke's arguments on the topic of natural religion are some of the best and most widely-known arguments in the period for the general deist position that natural philosophy in a broad sense grounds central doctrines of a universal religion.

Enlightenment deism first arises in England. In *On the Reasonableness of Christianity* (1695), Locke aims to establish the compatibility of reason and the teachings of Christianity. Though Locke himself is (like Newton, like Clarke) not a deist, the major English deists who follow (John Toland, *Christianity Not Mysterious* [1696]); Anthony Collins, *A Discourse of Freethinking* [1713]; Matthew Tindal, *Christianity as Old as Creation* [1730]) are influenced by Locke's work. Voltaire carries deism across the channel to France and advocates for it there over his long literary career. Toward the end-stage, the farcical stage, of the French Revolution, Robespierre institutes a form of deism, the so-called "Cult of the Supreme Being", as the official religion of the French state. Deism plays a role in the founding of the American republic as well. Many of the founding fathers (Jefferson, Franklin, Madison, Paine) author statements or tracts that are sympathetic to deism; and their deistic sympathies influence the place given (or not given) to religion in the new American state that they found.

Religion of the Heart. Opposition to deism derives sometimes from the perception of it as coldly rationalistic. The God of the deists, arrived at through a priori or empirical argument and referred to as the Prime Mover or Original Architect, is often perceived as distant and unconcerned with the daily struggles of human existence, and thus as not answering the human needs from which religion springs in the first place. Some important thinkers of the Enlightenment – notably Shaftesbury and Rousseau – present religion as founded on natural human sentiments, rather than on the operations of the intellect. Rousseau has his Savoyard Vicar declare, in his Profession of Faith in *Emile* (1762), that the idea of worshiping a beneficent deity arose in him initially as he reflected on his own situation in nature and his "heart began to glow with a sense of gratitude towards the author of our being". The Savoyard Vicar continues: "I adore the supreme power, and melt into tenderness at his goodness. I have no need to be taught artificial forms of worship; the dictates of nature

are sufficient. Is it not a natural consequence of self-love to honor those who protect us, and to love such as do us good?" This "natural" religion – opposed to the "artificial" religions enforced in the institutions – is often classed as a form of deism. But it deserves separate mention, because of its grounding in natural human sentiments, rather than in reason or in metaphysical or natural scientific problems of cosmology.

Fideism. Deism or natural religion of various sorts tends to rely on the claim that reason or human experience supports the hypothesis that there is a supreme being who created or authored the world. In one of the most important philosophical texts on natural religion to appear during the Enlightenment, David Hume's *Dialogues Concerning Natural Religion* (published posthumously in 1779), this supposition is criticized relentlessly, incisively and in detail. Naturally, the critical, questioning attitude characteristic of the Enlightenment in general is directed against the arguments on which natural religion is based. In Part Nine of the *Dialogues,* Samuel Clarke's "argument a priori" (as defended by the character Demea) is dispatched fairly quickly, but with a battery of arguments. But Hume is mainly concerned in the *Dialogues* with the other major pillar of natural religion in the Enlightenment, the "empirical" argument, the teleological argument or the argument from design. Cleanthes, the character who advances the design argument in the dialogue, proceeds from the rule for empirical reasoning that like effects prove like causes. He reasons that, given the resemblance between nature, which displays in many respects a "curious adaptation of means to ends", and a man-made machine, we must infer the cause of nature to be an intelligence like ours, though greater in proportion as nature surpasses in perfection the products of human intelligence. Philo, the skeptical voice in the *Dialogues*, presses Cleanthes' argument on many fronts. He points out that the argument is only as strong as the similarity between nature or parts of nature and man-made machines, and further, that a close scrutiny reveals that analogy to be weak. Moreover, according to the principle of

the argument, the stronger the evidence for an author (or authors) of nature, the *more like us* that author (or authors) should be taken to be. Consequently, according to Philo, the argument does not support the conclusion that *God* exists, taking God to be unitary, infinite, perfect, et cetera. Also, although the existence of evil and disorder in nature may serve actually to strengthen the case for the argument, given the disorder in human creations as well, the notion that God authors evil and disorder is disturbing. If one denies that there is disorder and evil in nature, however implausibly, the effect is to emphasize again the dissimilarity between nature and human products and thus weaken the central basis of the argument. With these and other considerations, Philo puts the proponent of the empirical argument in a difficult dialectical position. But Cleanthes is not moved. He holds the inference from the phenomenon of the curious adaptation of means to ends in nature to the existence of an intelligent and beneficent author to be so natural as to be impervious to the philosophical cavils raised by Philo. And, in the ambiguous conclusion of the work, Philo seems to agree. Though Hume himself seems to have been an atheist, one natural way to take the upshot of his *Dialogues* is that religious belief is so "natural" to us that rational criticism cannot unseat it. The ambiguous upshot of the work can be taken to be the impotence of rational criticism in the face of religious belief, rather than the illegitimacy of religious belief in the face of rational criticism. This tends toward fideism, the view according to which religious faith maintains its truth over against philosophical reasoning, which opposes but cannot defeat it. Fideism is most often associated with thinkers whose beliefs run contrary to the trends of the Enlightenment (Blaise Pascal, Johann-Georg Hamann, Søren Kierkegaard), but the skeptical strain in the Enlightenment, from Pierre Bayle through David Hume, expresses itself not only in atheism, but also in fideism.

Atheism. Atheism is more present in the French Enlightenment than elsewhere. In the writings of Denis Diderot, atheism is partly supported by

an expansive, dynamic conception of nature. According to the viewpoint developed by Diderot, we ought to search for the principles of natural order within natural processes themselves, not in a supernatural being. Even if we don't yet know the internal principles for the ordering and development of natural forms, the appeal to a transcendent author of such things is reminiscent, to Diderot's ear, of the appeal to Aristotelian "substantial forms" that was expressly rejected at the beginning of modern science as explaining nothing. The appeal to a transcendent author does not extend our understanding, but merely marks and fixes the limits of it. Atheism (combined with materialism) in the French Enlightenment is perhaps most identified with the Baron d'Holbach, whose *System of Nature* (1770) generated a great deal of controversy at the time for urging the case for atheism explicitly and emphatically. D'Holbach's system of nature is strongly influenced by Diderot's writings, though it displays less subtlety and dialectical sophistication. Though most Enlightenment thinkers hold that morality requires religion, in the sense that morality requires belief in a transcendent law-giver and in an after-life, d'Holbach (influenced in this respect by Spinoza, among others) makes the case for an ethical naturalism, an ethics that is free of any reference to a supernatural grounding or aspiration. Like Helvétius before him, d'Holbach presents an ethics in which virtue consists in enlightened self-interest. The metaphysical background of the ethics he presents is deterministic materialism. The Prussian enlightened despot, Frederick the Great, famously criticizes d'Holbach's book for exemplifying the incoherence that troubles the Enlightenment generally: while d'Holbach provides passionate moral critiques of existing religious and social and political institutions and practices, his own materialist, determinist conception of nature allows no place for moral "oughts" and prescriptions and values.

3. The Beautiful: Aesthetics in the Enlightenment

Modern systematic philosophical aesthetics not only first emerges in the context of the Enlightenment, but also flowers brilliantly there. As Ernst Cassirer notes, the eighteenth century not only thinks of itself as the "century of philosophy", but also as "the age of criticism," where criticism is centrally (though not only) art and literary criticism (Cassirer 1932, 255). Philosophical aesthetics flourishes in the period because of its strong affinities with the tendencies of the age. Alexander Baumgarten, the German philosopher in the school of Christian Wolff, founds systematic aesthetics in the period, in part through giving it its name. "Aesthetics" is derived from the Greek word for "senses", because for Baumgarten a science of the beautiful would be a science of the sensible, a science of sensible cognition. The Enlightenment in general re-discovers the value of the senses, not only in cognition, but in human lives in general, and so, given the intimate connection between beauty and human sensibility, the Enlightenment is naturally particularly interested in aesthetics. Also, the Enlightenment includes a general recovery and affirmation of the value of pleasure in human lives, against the tradition of Christian asceticism, and the flourishing of the arts, of the criticism of the arts and of the philosophical theorizing about beauty, promotes and is promoted by this recovery and affirmation. The Enlightenment also enthusiastically embraces the discovery and disclosure of rational order in nature, as manifest most clearly in the development of the new science. It seems to many theorists in the Enlightenment that the faculty of taste, the faculty by which we discern beauty, reveals to us some part of this order, a distinctive harmony, unities amidst variety. Thus, in the phenomenon of aesthetic pleasure, human sensibility discloses to us rational order, thus binding together two enthusiasms of the Enlightenment.

3.1 French Classicism and German Rationalism

In the early Enlightenment, especially in France, the emphasis is upon the discernment of an objective rational order, rather than upon the subject's

sensual aesthetic pleasure. Though Descartes' philosophical system does not include a theory of taste or of beauty, his mathematical model of the physical universe inspires the aesthetics of French classicism. French classicism begins from the classical maxim that the beautiful is the true. Nicolas Boileau writes in his influential didactic poem, *The Art of Poetry* (1674), in which he lays down rules for good versification within different genres, that "Nothing is beautiful but the true, the true alone is lovable." In the period the true is conceived of as an objective rational order. According to the classical conception of art that dominates in the period, art *imitates* nature, though not nature as given in disordered experience, but the *ideal* nature, the ideal in which we can discern and enjoy "unity in multiplicity." In French classicism, aesthetics is very much under the influence of, and indeed modeled on, systematic, rigorous theoretical science of nature. Just as in Descartes' model of science, where knowledge of all particulars depends on prior knowledge of the principle from which the particulars are deduced, so also in the aesthetics of French classicism, the demand is for systematization under a single, universal principle. The subjection of artistic phenomena to universal rules and principles is expressed, for example, in the title of Charles Batteaux's main work, *The Fine Arts Reduced to a Single Principle* (1746), as well as in Boileau's rules for good versification.

In Germany in the eighteenth century, Christian Wolff's systematic rationalist metaphysics forms the basis for much of the reflection on aesthetics, though sometimes as a set of doctrines to be argued against. Wolff affirms the classical dictum that beauty is truth; beauty is truth perceived through the feeling of pleasure. Wolff understands beauty to consist in the perfection in things, which he understands in turn to consist in a harmony or order of a manifold. We judge something beautiful through a feeling of pleasure when we sense in it this harmony or perfection. Beauty is, for Wolff, the sensitive cognition of perfection. Thus, for Wolff, beauty corresponds to objective features of the world, but

judgments of beauty are relative to us also, insofar as they are based on the human faculty of sensibility.

3.2 Empiricism and Subjectivism

Though philosophical rationalism forms the basis of aesthetics in the early Enlightenment in France and Germany, thinkers in the empiricist tradition in England and Scotland introduce many of the salient themes of Enlightenment aesthetics. In particular, with the rise of empiricism and subjectivism in this domain, attention shifts to the ground and nature of the subject's experience of beauty, the subject's aesthetic response. Lord Shaftesbury, though not himself an empiricist or subjectivist in aesthetics, makes significant contributions to this development. Shaftesbury re-iterates the classical equation, "all beauty is truth," but the truth that beauty is for Shaftesbury is not an objective rational order that could also be known conceptually. Though beauty is, for Shaftesbury, a kind of harmony that is independent of the human mind, under the influence of Plotinus, he understands the human being's immediate intuition of the beautiful as a kind of participation in the original harmony. Shaftesbury focuses attention on the nature of the subject's response to beauty, as elevating the person, also morally. He maintains that aesthetic response consists in a *disinterested* unegoistic pleasure; the discovery of this capacity for disinterested pleasure in harmony shows the way for the development of his ethics that has a similar grounding. And, in fact, in seeing aesthetic response as elevating oneself above self-interested pursuits, through cultivating one's receptivity to disinterested pleasure, Shaftesbury ties tightly together aesthetics and ethics, morality and beauty, and in that respect also contributes to a trend of the period. Also, in placing the emphasis on the subject's response to beauty, rather than on the objective characteristics of the beautiful, Shaftesbury makes aesthetics belong to the general Enlightenment interest in human nature. Thinkers of

the period find in our receptivity to beauty a key both to understanding both distinctively human nature and its perfection.

Francis Hutcheson follows Shaftesbury in his emphasis on the subject's aesthetic response, on the distinctive sort of pleasure that the beautiful elicits in us. Partly because the Neo-Platonic influence, so pronounced in Shaftesbury's aesthetics, is washed out of Hutcheson's, to be replaced by a more thorough-going empiricism, Hutcheson understands this distinctive aesthetic pleasure as more akin to a secondary quality. Thus, Hutcheson's aesthetic work raises the prominent question whether "beauty" refers to something objective at all or whether beauty is "nothing more" than a human idea or experience. As in the domain of Enlightenment ethics, so with Enlightenment aesthetics too, the step from Shaftesbury to Hutcheson marks a step toward subjectivism. Hutcheson writes in one of his *Two Treatises*, his *Inquiry Concerning Beauty, Order, Harmony, Design* (1725) that "the word 'beauty' is taken for the *idea raised in us*, and a *sense* of beauty for *our power of receiving this idea*" (Section I, Article IX). However, though Hutcheson understands beauty to be an idea in us, he takes this idea to be "excited" or "occasioned" in us by distinctive objective qualities, in particular by objects that display "*uniformity amidst variety*" (ibid., Section II, Article III). In the very title of Hutcheson's work above, we see the importance of the classical ideas of (rational) order and harmony in Hutcheson's aesthetic theory, even as he sets the tenor for much Enlightenment discussion of aesthetics through placing the emphasis on the subjective idea and aesthetic response.

David Hume's famous essay on "the standard of taste" raises and addresses the epistemological problem raised by subjectivism in aesthetics. If beauty is an idea in us, rather than a feature of objects independent of us, then how do we understand the possibility of correctness and incorrectness – how do we understand the possibility of standards of judgment – in this domain? The problem is posed more

clearly for Hume because he intensifies Hutcheson's subjectivism. He writes in the *Treatise* that "pleasure and pain….are not only necessary attendants of beauty and deformity, but constitute their very essence" (*Treatise*, Book II, part I, section viii). But if a judgment of taste is based on, or expresses, subjective sentiments, how can it be incorrect? In his response to this question, Hume accounts for the expectation of agreement in judgments of taste by appealing to the fact that we share a common human nature, and he accounts for 'objectivity' or expertise in judgments of taste, within the context of his subjectivism, by appealing to the normative responses of well-placed observers. Both of these points (the commonality of human nature and the securing of 'objectivity' in judgments based on sentiments by appeal to the normative responses of appropriately placed observers) are typical of the period more generally, and especially of the strong empiricist strain in the Enlightenment. Hume develops the empiricist line in aesthetics to the point where little remains of the classical emphasis on the order or harmony or truth that is, according to the French classicists, apprehended and appreciated in our aesthetic responses to the beautiful, and thus, according to the classicists, the ground of aesthetic responses.

3.3 Late Enlightenment Aesthetics

Immanuel Kant faces squarely the problem of the normativity of judgments of taste. Influenced by Hutcheson and the British empiricist tradition in general, Kant understands judgments of taste to be founded on a distinctive sort of feeling, a *disinterested* pleasure. In taking judgments of taste to be subjective (they are founded on the subject's feeling of pleasure) and non-cognitive (such judgments do not subsume representations under concepts and thus do not ascribe properties to objects), Kant breaks with the German rationalist school. However Kant continues to maintain that judgments of beauty are like cognitive judgments in making a legitimate claim to universal agreement – in

contrast to judgments of the agreeable. The question is how to vindicate the legitimacy of this demand. Kant argues that the distinctive pleasure underlying judgments of taste is the experience of the harmony of the faculties of the imagination and the understanding, a harmony that arises through their "free play" in the process of cognizing objects on the basis of given sensible intuition. The harmony is "free" in an experience of beauty in the sense that it is not forced by rules of the understanding, as is the agreement among the faculties in acts of cognition. The order and harmony that we experience in the face of the beautiful is subjective, according to Kant; but it is at the same time universal and normative, by virtue of its relation to the conditions of human cognition.

The emphasis Kant places on the role of the activity of the imagination in aesthetic pleasure and discernment typifies a trend in Enlightenment thought. Whereas early in the Enlightenment, in French classicism, and to some extent in Christian Wolff and other figures of German rationalism, the emphasis is on the more-or-less static rational order and proportion and on rigid universal rules or laws of reason, the trend during the development of Enlightenment aesthetics is toward emphasis on the *play* of the imagination and its fecundity in generating associations.

Denis Diderot is an important and influential author on aesthetics. He wrote the entry "On the Origin and Nature of the Beautiful" for the *Encyclopedia* (1752). Like Lessing in Germany, Diderot not only philosophized about art and beauty, but also wrote plays and influential art criticism. Diderot is strongly influenced in his writings on aesthetics by the empiricism in England and Scotland, but his writing is not limited to that standpoint. Diderot repeats the classical dictum that art should imitate nature, but, whereas, for French classicists, the nature that art should imitate is *ideal* nature – a static, universal rational order – for Diderot, nature is dynamic and productive. For Diderot, the nature the artist ought to imitate is the *real* nature we experience, warts and all (as it were). The

particularism and realism of Diderot's aesthetics is based on a critique of the standpoint of French classicism (see Cassirer 1935, p. 295f.). This critique exposes the artistic rules represented by French classicists as universal rules of reason as nothing more than *conventions* marking what is considered *proper* within a certain tradition. In other words, the prescriptions within the French classical tradition are *artificial*, not *natural*, and constitute fetters to artistic genius. Diderot takes liberation from such fetters to come from turning to the task of observing and imitating *actual nature*. Diderot's emphasis on the primeval productive power and abundance of nature in his aesthetic writings contributes to the trend toward focus on artistic creation and expression (as opposed to artistic appreciation and discernment) that is a characteristic of the late Enlightenment and the transition to Romanticism.

Lessing's aesthetic writings play an important role in elevating the aesthetic category of expressiveness. In his famous *Laocoön: An Essay on the Limits of Painting and Poetry* (1766), Lessing argues, by comparing the famous Greek statue with the representation of Laocoön's suffering in Virgil's poetry, that the aims of poetry and of the visual arts are not identical; he argues that the aim of poetry is not beauty, but expression. In elevating the aesthetic category of expressiveness, Lessing challenges the notion that all art is imitation of nature. His argument also challenges the notion that all the various arts can be deduced from a single principle. Lessing's argument in *Laocoön* supports the contrary thesis that the distinct arts have distinct aims and methods, and that each should be understood on its own terms, not in terms of an abstract general principle from which all arts are to be deduced. For some, especially for critics of the Enlightenment, in this point Lessing is already beyond the Enlightenment. Certainly it is true that the emphasis on the individual or particular, over against the universal, which one finds in other late Enlightenment thinkers, is in tension with Enlightenment tenets. Herder (following Hamann to some extent) argues that each individual art *object*

has to be understood in its own terms, as a totality complete unto itself. With Herder's stark emphasis on individuality in aesthetics, over against universality, the supplanting of the Enlightenment with Romanticism and Historicism is well advanced. But, according to the point of view taken in this entry, the conception of the Enlightenment according to which it is distinguished by its prioritization of the order of abstract, universal laws and principles, over against concrete particulars and the differences amongst them, is too narrow; it fails to account for much of the characteristic richness in the thought of the period. Indeed aesthetics itself, as a discipline, which, as noted, is founded in the Enlightenment by the German rationalist, Alexander Baumgarten, owes its existence to the tendency in the Enlightenment to search for and discover distinct laws for distinct kinds of phenomena (as opposed to insisting that all phenomena be made intelligible through the same set of general laws and principles). Baumgarten founds aesthetics as a 'science' through the attempt to establish the sensible domain as cognizable in a way different from that which prevails in metaphysics. Aesthetics in Germany in the eighteenth century, from Wolff to Herder, both typifies many of the trends of the Enlightenment and marks the field where the Enlightenment yields to competing worldviews.

Bibliography

Primary Literature

Bacon, F., 1620. *The New Organon (Novum Organum)*, ed. by Lisa Jardine and Michael Silverthorne, Cambridge: Cambridge University Press, 2000.

Batteaux, C., 1746. *Les beaux arts réduit à un même principe* (*The Fine Arts Reduced to a Single Principle*). Paris: Chez Durand.

Bayle, P., 1697, *Historical and Critical Dictionary*, 2nd edition, 1702, tr. by R. Popkin, Indianapolis: Bobbs Merrill, 1965.

Rationalism vs. Empiricism

First published Thu Aug 19, 2004; substantive revision Thu Jul 6, 2017

The dispute between rationalism and empiricism concerns the extent to which we are dependent upon sense experience in our effort to gain knowledge. Rationalists claim that there are significant ways in which our concepts and knowledge are gained independently of sense experience. Empiricists claim that sense experience is the ultimate source of all our concepts and knowledge.

Rationalists generally develop their view in two ways. First, they argue that there are cases where the content of our concepts or knowledge outstrips the information that sense experience can provide. Second, they construct accounts of how reason in some form or other provides that additional information about the world. Empiricists present complementary lines of thought. First, they develop accounts of how experience provides the information that rationalists cite, insofar as we have it in the first place. (Empiricists will at times opt for skepticism as an alternative to rationalism: if experience cannot provide the concepts or knowledge the rationalists cite, then we don't have them.) Second, empiricists attack the rationalists' accounts of how reason is a source of concepts or knowledge.

1. Introduction

The dispute between rationalism and empiricism takes place within epistemology, the branch of philosophy devoted to studying the nature, sources and limits of knowledge. The defining questions of epistemology include the following.

1. What is the nature of propositional knowledge, knowledge that a particular proposition about the world is true?

   To know a proposition, we must believe it and it must be true, but something more is required, something that distinguishes knowledge from a lucky guess. Let's call this additional element 'warrant'. A good deal of philosophical work has been invested in trying to determine the nature of warrant.

2. How can we gain knowledge?

   We can form true beliefs just by making lucky guesses. How to gain warranted beliefs is less clear. Moreover, to know the world, we must think about it, and it is unclear how we gain the concepts we use in thought or what assurance, if any, we have that the ways in which we divide up the world using our concepts correspond to divisions that actually exist.

3. What are the limits of our knowledge?

Some aspects of the world may be within the limits of our thought but beyond the limits of our knowledge; faced with competing descriptions of them, we cannot know which description is true. Some aspects of the world may even be beyond the limits of our thought, so that we cannot form intelligible descriptions of them, let alone know that a particular description is true.

The disagreement between rationalists and empiricists primarily concerns the second question, regarding the sources of our concepts and knowledge. In some instances, their disagreement on this topic leads them to give conflicting responses to the other questions as well. They may disagree over the nature of warrant or about the limits of our thought and knowledge. Our focus here will be on the competing rationalist and empiricist responses to the second question.

1.1 Rationalism

To be a rationalist is to adopt at least one of three claims. The Intuition/Deduction thesis concerns how we become warranted in believing propositions in a particular subject area.

The Intuition/Deduction Thesis: Some propositions in a particular subject area, S, are knowable by us by intuition alone; still others are knowable by being deduced from intuited propositions.

Intuition is a form of rational insight. Intellectually grasping a proposition, we just "see" it to be true in such a way as to form a true, warranted belief in it. (As discussed in Section 2 below, the nature of this intellectual "seeing" needs explanation.) Deduction is a process in which we derive conclusions from intuited premises through valid arguments, ones in which the conclusion must be true if the premises are true. We intuit, for example, that the number three is prime and that it is greater than two. We then deduce from this knowledge that there is a prime number greater than

two. Intuition and deduction thus provide us with knowledge *a priori*, which is to say knowledge gained independently of sense experience.

We can generate different versions of the Intuition/Deduction thesis by substituting different subject areas for the variable 'S'. Some rationalists take mathematics to be knowable by intuition and deduction. Some place ethical truths in this category. Some include metaphysical claims, such as that God exists, we have free will, and our mind and body are distinct substances. The more propositions rationalists include within the range of intuition and deduction, and the more controversial the truth of those propositions or the claims to know them, the more radical their rationalism.

Rationalists also vary the strength of their view by adjusting their understanding of warrant. Some take warranted beliefs to be beyond even the slightest doubt and claim that intuition and deduction provide beliefs of this high epistemic status. Others interpret warrant more conservatively, say as belief beyond a reasonable doubt, and claim that intuition and deduction provide beliefs of that caliber. Still another dimension of rationalism depends on how its proponents understand the connection between intuition, on the one hand, and truth, on the other. Some take intuition to be infallible, claiming that whatever we intuit must be true. Others allow for the possibility of false intuited propositions.

The second thesis associated with rationalism is the Innate Knowledge thesis.

> *The Innate Knowledge Thesis*: We have knowledge of some truths in a particular subject area, S, as part of our rational nature.

Like the Intuition/Deduction thesis, the Innate Knowledge thesis asserts the existence of knowledge gained *a priori*, independently of experience. The difference between them rests in the accompanying understanding of

how this *a priori* knowledge is gained. The Intuition/Deduction thesis cites intuition and subsequent deductive reasoning. The Innate Knowledge thesis offers our rational nature. Our innate knowledge is not learned through either sense experience or intuition and deduction. It is just part of our nature. Experiences may trigger a process by which we bring this knowledge to consciousness, but the experiences do not provide us with the knowledge itself. It has in some way been with us all along. According to some rationalists, we gained the knowledge in an earlier existence. According to others, God provided us with it at creation. Still others say it is part of our nature through natural selection.

We get different versions of the Innate Knowledge thesis by substituting different subject areas for the variable 'S'. Once again, the more subjects included within the range of the thesis or the more controversial the claim to have knowledge in them, the more radical the form of rationalism. Stronger and weaker understandings of warrant yield stronger and weaker versions of the thesis as well.

The third important thesis of rationalism is the Innate Concept thesis.

> *The Innate Concept Thesis*: We have some of the concepts we employ in a particular subject area, S, as part of our rational nature.

According to the Innate Concept thesis, some of our concepts are not gained from experience. They are part of our rational nature in such a way that, while sense experiences may trigger a process by which they are brought to consciousness, experience does not provide the concepts or determine the information they contain. Some claim that the Innate Concept thesis is entailed by the Innate Knowledge Thesis; a particular instance of knowledge can only be innate if the concepts that are contained in the known proposition are also innate. This is Locke's position (1690, Book I, Chapter IV, Section 1, p. 91). Others, such as Carruthers, argue

against this connection (1992, pp. 53–54). The content and strength of the Innate Concept thesis varies with the concepts claimed to be innate. The more a concept seems removed from experience and the mental operations we can perform on experience the more plausibly it may be claimed to be innate. Since we do not experience perfect triangles but do experience pains, our concept of the former is a more promising candidate for being innate than our concept of the latter.

The Intuition/Deduction thesis, the Innate Knowledge thesis, and the Innate Concept thesis are essential to rationalism: to be a rationalist is to adopt at least one of them. Two other closely related theses are generally adopted by rationalists, although one can certainly be a rationalist without adopting either of them. The first is that experience cannot provide what we gain from reason.

> *The Indispensability of Reason Thesis*: The knowledge we gain in subject area, S, by intuition and deduction, as well as the ideas and instances of knowledge in S that are innate to us, could not have been gained by us through sense experience.

The second is that reason is superior to experience as a source of knowledge.

> *The Superiority of Reason Thesis*: The knowledge we gain in subject area S by intuition and deduction or have innately is superior to any knowledge gained by sense experience.

How reason is superior needs explanation, and rationalists have offered different accounts. One view, generally associated with Descartes (1628, Rules II and III, pp. 1–4), is that what we know *a priori* is certain, beyond even the slightest doubt, while what we believe, or even know, on the basis of sense experience is at least somewhat uncertain. Another view, generally associated with Plato (*Republic* 479e-484c), locates the

superiority of *a priori* knowledge in the objects known. What we know by reason alone, a Platonic form, say, is superior in an important metaphysical way, e.g. unchanging, eternal, perfect, a higher degree of being, to what we are aware of through sense experience.

Most forms of rationalism involve notable commitments to other philosophical positions. One is a commitment to the denial of scepticism for at least some area of knowledge. If we claim to know some truths by intuition or deduction or to have some innate knowledge, we obviously reject scepticism with regard to those truths. Rationalism in the form of the Intuition/Deduction thesis is also committed to epistemic foundationalism, the view that we know some truths without basing our belief in them on any others and that we then use this foundational knowledge to know more truths.

1.2 Empiricism

Empiricists endorse the following claim for some subject area.

> *The Empiricism Thesis*: We have no source of knowledge in S or for the concepts we use in S other than sense experience.

Empiricism about a particular subject rejects the corresponding version of the Intuition/Deduction thesis and Innate Knowledge thesis. Insofar as we have knowledge in the subject, our knowledge is *a posteriori*, dependent upon sense experience. Empiricists also deny the implication of the corresponding Innate Concept thesis that we have innate ideas in the subject area. Sense experience is our only source of ideas. They reject the corresponding version of the Superiority of Reason thesis. Since reason alone does not give us any knowledge, it certainly does not give us superior knowledge. Empiricists generally reject the Indispensability of Reason thesis, though they need not. The Empiricism thesis does not entail that we have empirical knowledge. It entails that knowledge can only be

gained, *if at all*, by experience. Empiricists may assert, as some do for some subjects, that the rationalists are correct to claim that experience cannot give us knowledge. The conclusion they draw from this rationalist lesson is that we do not know at all.

I have stated the basic claims of rationalism and empiricism so that each is relative to a particular subject area. Rationalism and empiricism, so relativized, need not conflict. We can be rationalists in mathematics or a particular area of mathematics and empiricists in all or some of the physical sciences. Rationalism and empiricism only conflict when formulated to cover the same subject. Then the debate, Rationalism vs. Empiricism, is joined. The fact that philosophers can be both rationalists and empiricists has implications for the classification schemes often employed in the history of philosophy, especially the one traditionally used to describe the Early Modern Period of the seventeenth and eighteenth centuries leading up to Kant. It is standard practice to group the major philosophers of this period as either rationalists or empiricists and to suggest that those under one heading share a common agenda in opposition to those under the other. Thus, Descartes, Spinoza and Leibniz are the Continental Rationalists in opposition to Locke, Berkeley and Hume, the British Empiricists. We should adopt such general classification schemes with caution. The views of the individual philosophers are more subtle and complex than the simple-minded classification suggests. (See Loeb (1981) and Kenny (1986) for important discussions of this point.) Locke rejects rationalism in the form of any version of the Innate Knowledge or Innate Concept theses, but he nonetheless adopts the Intuition/Deduction thesis with regard to our knowledge of God's existence. Descartes and Locke have remarkably similar views on the nature of our ideas, even though Descartes takes many to be innate, while Locke ties them all to experience. The rationalist/empiricist classification also encourages us to expect the philosophers on each side of the divide to have common research programs in areas beyond epistemology. Thus,

Descartes, Spinoza and Leibniz are mistakenly seen as applying a reason-centered epistemology to a common metaphysical agenda, with each trying to improve on the efforts of the one before, while Locke, Berkeley and Hume are mistakenly seen as gradually rejecting those metaphysical claims, with each consciously trying to improve on the efforts of his predecessors. It is also important to note that the rationalist/empiricist distinction is not exhaustive of the possible sources of knowledge. One might claim, for example, that we can gain knowledge in a particular area by a form of Divine revelation or insight that is a product of neither reason nor sense experience. In short, when used carelessly, the labels 'rationalist' and 'empiricist,' as well as the slogan that is the title of this essay, 'Rationalism vs. Empiricism,' can retard rather than advance our understanding.

Nonetheless, an important debate properly described as 'Rationalism vs. Empiricism' is joined whenever the claims for each view are formulated to cover the same subject. What is perhaps the most interesting form of the debate occurs when we take the relevant subject to be truths about the external world, the world beyond our own minds. A full-fledged rationalist with regard to our knowledge of the external world holds that some external world truths can and must be known *a priori*, that some of the ideas required for that knowledge are and must be innate, and that this knowledge is superior to any that experience could ever provide. The full-fledged empiricist about our knowledge of the external world replies that, when it comes to the nature of the world beyond our own minds, experience is our sole source of information. Reason might inform us of the relations among our ideas, but those ideas themselves can only be gained, and any truths about the external reality they represent can only be known, on the basis of sense experience. This debate concerning our knowledge of the external world will generally be our main focus in what follows.

Historically, the rationalist/empiricist dispute in epistemology has extended into the area of metaphysics, where philosophers are concerned with the basic nature of reality, including the existence of God and such aspects of our nature as freewill and the relation between the mind and body. Major rationalists (e.g., Descartes 1641) have presented metaphysical theories, which they have claimed to know by reason alone. Major empiricists (e.g., Hume 1739–40) have rejected the theories as either speculation, beyond what we can learn from experience, or nonsensical attempts to describe aspects of the world beyond the concepts experience can provide. The debate raises the issue of metaphysics as an area of knowledge. Kant puts the driving assumption clearly:

> The very concept of metaphysics ensures that the sources of metaphysics can't be empirical. If something could be known through the senses, that would automatically show that it doesn't belong to metaphysics; that's an upshot of the meaning of the word 'metaphysics.' Its basic principles can never be taken from experience, nor can its basic concepts; for it is not to be physical but metaphysical knowledge, so it must be beyond experience. (1783, Preamble, I, p. 7)

The possibility then of metaphysics so understood, as an area of human knowledge, hinges on how we resolve the rationalist/empiricist debate. The debate also extends into ethics. Some moral objectivists (e.g., Ross 1930 and Huemer 2005) take us to know some fundamental objective moral truths by intuition, while some moral skeptics, who reject such knowledge, (e.g., Mackie 1977) find the appeal to a faculty of moral intuition utterly implausible. More recently, the rationalist/empiricist debate has extended to discussions (e.g., Bealer 1999 and Alexander & Weinberg 2007) of the very nature of philosophical inquiry: to what extent are philosophical questions to be answered by appeals to reason or experience?

2. The Intuition/Deduction Thesis

The Intuition/Deduction thesis claims that we can know some propositions by intuition and still more by deduction. Many empiricists (e.g., Hume 1748) have been willing to accept the thesis so long as it is restricted to propositions solely about the relations among our own concepts. We can, they agree, know by intuition that our concept of God includes our concept of omniscience. Just by examining the concepts, we can intellectually grasp that the one includes the other. The debate between rationalists and empiricists is joined when the former assert, and the latter deny, the Intuition/Deduction thesis with regard to propositions that contain substantive information about the external world. Rationalists, such as Descartes, have claimed that we can know by intuition and deduction that God exists and created the world, that our mind and body are distinct substances, and that the angles of a triangle equal two right angles, where all of these claims are truths about an external reality independent of our thought. Such substantive versions of the Intuition/Deduction thesis are our concern in this section.

One defense of the Intuition/Deduction thesis assumes that we know some substantive external world truths, adds an analysis of what knowledge requires, and concludes that our knowledge must result from intuition and deduction. Descartes claims that knowledge requires certainty and that certainty about the external world is beyond what empirical evidence can provide. We can never be sure our sensory impressions are not part of a dream or a massive, demon orchestrated, deception. Only intuition and deduction can provide the certainty needed for knowledge, and, given that we have some substantive knowledge of the external world, the Intuition/Deduction thesis is true. As Descartes tells us, "all knowledge is certain and evident cognition" (1628, Rule II, p. 1) and when we "review all the actions of the intellect by means of which we are able to arrive at a

knowledge of things with no fear of being mistaken," we "recognize only two: intuition and deduction" (1628, Rule III, p. 3).

This line of argument is one of the least compelling in the rationalist arsenal. First, the assumption that knowledge requires certainty comes at a heavy cost, as it rules out so much of what we commonly take ourselves to know. Second, as many contemporary rationalists accept, intuition is not always a source of certain knowledge. The possibility of a deceiver gives us a reason to doubt our intuitions as well as our empirical beliefs. For all we know, a deceiver might cause us to intuit false propositions, just as one might cause us to have perceptions of nonexistent objects. Descartes's classic way of meeting this challenge in the *Meditations* is to argue that we can know with certainty that no such deceiver interferes with our intuitions and deductions. They are infallible, as God guarantees their truth. The problem, known as the Cartesian Circle, is that Descartes's account of how we gain this knowledge begs the question, by attempting to deduce the conclusion that all our intuitions are true from intuited premises. Moreover, his account does not touch a remaining problem that he himself notes (1628, Rule VII, p. 7): Deductions of any appreciable length rely on our fallible memory.

A more plausible argument for the Intuition/Deduction thesis again assumes that we know some particular, external world truths, and then appeals to the nature of what we know, rather than to the nature of knowledge itself, to argue that our knowledge must result from intuition and deduction. Leibniz (1704) tells us the following.

> The senses, although they are necessary for all our actual knowledge, are not sufficient to give us the whole of it, since the senses never give anything but instances, that is to say particular or individual truths. Now all the instances which confirm a general truth, however numerous they may be, are not sufficient to

establish the universal necessity of this same truth, for it does not follow that what happened before will happen in the same way again. … From which it appears that necessary truths, such as we find in pure mathematics, and particularly in arithmetic and geometry, must have principles whose proof does not depend on instances, nor consequently on the testimony of the senses, although without the senses it would never have occurred to us to think of them… (1704, Preface, pp. 150–151)

Leibniz goes on to describe our mathematical knowledge as "innate," and his argument may be directed to support the Innate Knowledge thesis rather than the Intuition/Deduction thesis. For our purposes here, we can relate it to the latter, however: We have substantive knowledge about the external world in mathematics, and what we know in that area, we know to be necessarily true. Experience cannot warrant beliefs about what is necessarily the case. Hence, experience cannot be the source of our knowledge. The best explanation of our knowledge is that we gain it by intuition and deduction. Leibniz mentions logic, metaphysics and morals as other areas in which our knowledge similarly outstrips what experience can provide. Judgments in logic and metaphysics involve forms of necessity beyond what experience can support. Judgments in morals involve a form of obligation or value that lies beyond experience, which only informs us about what is the case rather than about what ought to be.

The strength of this argument varies with its examples of purported knowledge. Insofar as we focus on controversial claims in metaphysics, e.g., that God exists, that our mind is a distinct substance from our body, the initial premise that we know the claims is less than compelling. Taken with regard to other areas, however, the argument clearly has legs. We know a great deal of mathematics, and what we know, we know to be necessarily true. None of our experiences warrants a belief in such necessity, and we do not seem to base our knowledge on any experiences.

The warrant that provides us with knowledge arises from an intellectual grasp of the propositions which is clearly part of our learning. Similarly, we seem to have such moral knowledge as that, all other things being equal, it is wrong to break a promise and that pleasure is intrinsically good. No empirical lesson about how things are can warrant such knowledge of how they ought to be.

This argument for the Intuition/Deduction thesis raises additional questions which rationalists must answer. Insofar as they maintain that our knowledge of necessary truths in mathematics or elsewhere by intuition and deduction is substantive knowledge of the external world, they owe us an account of this form of necessity. Many empiricists stand ready to argue that "necessity resides in the way we talk about things, not in the things we talk about" (Quine 1966, p. 174). Similarly, if rationalists claim that our knowledge in morals is knowledge of an objective form of obligation, they owe us an account of how objective values are part of a world of apparently valueless facts.

Perhaps most of all, rationalist defenders of the Intuition/Deduction thesis owe us an account of what intuition is and how it provides warranted true beliefs about the external world. What is it to intuit a proposition and how does that act of intuition support a warranted belief? Their argument presents intuition and deduction as an explanation of assumed knowledge that can't—they say—be explained by experience, but such an explanation by intuition and deduction requires that we have a clear understanding of intuition and how it supports warranted beliefs. Metaphorical characterizations of intuition as intellectual "grasping" or "seeing" are not enough, and if intuition is some form of intellectual "grasping," it appears that all that is grasped is relations among our concepts, rather than facts about the external world. One current approach to the issue involves an appeal to Phenomenal Conservatism (Huemer 2001), the principle that if it seems to one as if something is the case, then one is prima facie justified in

believing that it is so. Intuitions are then taken to be a particular sort of seeming or appearance: "[A]n intuition that p is a state of its seeming to one that p that is not dependent on inference from other beliefs and that results from thinking about p, as opposed to perceiving, remembering, or introspecting" (Hummer 2005, p. 102). Just as it can visually seem or appear to one as if there's a tree outside the window, it can intellectually seem or appear to one as if nothing can be both entirely red and entirely green. This approach aims to demystify intuitions; they are but one more form of seeming-state along with ones we gain from sense perception, memory and introspection. It does not, however, tell us all we need to know. Any intellectual faculty, whether it be sense perception, memory, introspection or intuition, provides us with warranted beliefs only if it is generally reliable. The reliability of sense perception stems from the causal connection between how external objects are and how we experience them. What accounts for the reliability of our intuitions regarding the external world? Is our intuition of a particular true proposition the outcome of some causal interaction between ourselves and some aspect of the world? What aspect? What is the nature of this causal interaction? That the number three is prime does not appear to cause anything, let alone our intuition that it is prime. As Michael Huemer (2005, p. 123) points out in mounting his own defense of moral intuitionism, "The challenge for the moral realist, then, is to explain how it would be anything more than chance if my moral beliefs were true, given that I do not interact with moral properties."

These issues are made all the more pressing by the classic empiricist response to the argument. The reply is generally credited to Hume and begins with a division of all true propositions into two categories.

> All the objects of human reason or inquiry may naturally be divided into two kinds, to wit, "Relations of Ideas," and "Matters of Fact." Of the first are the sciences of Geometry, Algebra, and

Arithmetic, and, in short, every affirmation which is either intuitively or demonstratively certain. That the square of the hypotenuse is equal to the square of the two sides is a proposition which expresses a relation between these figures. That three times five is equal to half of thirty expresses a relation between these numbers. Propositions of this kind are discoverable by the mere operation of thought, without dependence on what is anywhere existent in the universe. Though there never were a circle or triangle in nature, the truths demonstrated by Euclid would forever retain their certainty and evidence. Matters of fact, which are the second objects of human reason, are not ascertained in the same manner, nor is our evidence of their truth, however great, of a like nature with the foregoing. The contrary of every matter of fact is still possible, because it can never imply a contradiction and is conceived by the mind with the same facility and distinctness as if ever so conformable to reality. (Hume 1748, Section IV, Part 1, p. 40)

Intuition and deduction can provide us with knowledge of necessary truths such as those found in mathematics and logic, but such knowledge is not substantive knowledge of the external world. It is only knowledge of the relations of our own ideas. If the rationalist shifts the argument so it appeals to knowledge in morals, Hume's reply is to offer an analysis of our moral concepts by which such knowledge is empirically gained knowledge of matters of fact.

Morals and criticism are not so properly objects of the understanding as of taste and sentiment. Beauty, whether moral or natural, is felt more properly than perceived. Or if we reason concerning it and endeavor to fix the standard, we regard a new fact, to wit, the general taste of mankind, or some other fact which may be the object of reasoning and inquiry. (Hume 1748, Section

XII, Part 3, p. 173)

If the rationalist appeals to our knowledge in metaphysics to support the argument, Hume denies that we have such knowledge.

> If we take in our hand any volume--of divinity or school metaphysics, for instance--let us ask, Does it contain any abstract reasoning concerning quantity or number? No. Does it contain any experimental reasoning concerning matter of fact and existence? No. Commit it then to the flames, for it can contain nothing but sophistry and illusion. (Hume 1748, Section XII, Part 3, p. 173)

An updated version of this general empiricist reply, with an increased emphasis on language and the nature of meaning, is given in the twentieth-century by A. J. Ayer's version of logical positivism. Adopting positivism's verification theory of meaning, Ayer assigns every cognitively meaningful sentence to one of two categories: either it is a tautology, and so true solely by virtue of the meaning of its terms and provides no substantive information about the world, or it is open to empirical verification. There is, then, no room for knowledge about the external world by intuition or deduction.

> There can be no *a priori* knowledge of reality. For … the truths of pure reason, the propositions which we know to be valid independently of all experience, are so only in virtue of their lack of factual content … [By contrast] empirical propositions are one and all hypotheses which may be confirmed or discredited in actual sense experience. [Ayer 1952, pp. 86; 93–94]

The rationalists' argument for the Intuition/Deduction thesis goes wrong at the start, according to empiricists, by assuming that we can have substantive knowledge of the external world that outstrips what experience can warrant. We cannot.

This empiricist reply faces challenges of its own. Our knowledge of mathematics seems to be about something more than our own concepts. Our knowledge of moral judgments seems to concern not just how we feel or act but how we ought to behave. The general principles that provide a basis for the empiricist view, e.g. Hume's overall account of our ideas, the Verification Principle of Meaning, are problematic in their own right. In various formulations, the Verification Principle fails its own test for having cognitive meaning. A careful analysis of Hume's *Inquiry,* relative to its own principles, may require us to consign large sections of it to the flames.

In all, rationalists have a strong argument for the Intuition/Deduction thesis relative to our substantive knowledge of the external world, but its success rests on how well they can answer questions about the nature and epistemic force of intuition made all the more pressing by the classic empiricist reply.

3. The Innate Knowledge Thesis

The Innate Knowledge thesis joins the Intuition/Deduction thesis in asserting that we have *a priori* knowledge, but it does not offer intuition and deduction as the source of that knowledge. It takes our *a priori* knowledge to be part of our rational nature. Experience may trigger our awareness of this knowledge, but it does not provide us with it. The knowledge is already there.

Plato presents an early version of the Innate Knowledge thesis in the *Meno* as the doctrine of knowledge by recollection. The doctrine is motivated in part by a paradox that arises when we attempt to explain the nature of inquiry. How do we gain knowledge of a theorem in geometry? We inquire into the matter. Yet, knowledge by inquiry seems impossible (*Meno*, 80d-e). We either already know the theorem at the start of our investigation or we do not. If we already have the knowledge, there is no place for inquiry.

If we lack the knowledge, we don't know what we are seeking and cannot recognize it when we find it. Either way we cannot gain knowledge of the theorem by inquiry. Yet, we do know some theorems.

The doctrine of knowledge by recollection offers a solution. When we inquire into the truth of a theorem, we both do and do not already know it. We have knowledge in the form of a memory gained from our soul's knowledge of the theorem prior to its union with our body. We lack knowledge in that, in our soul's unification with the body, it has forgotten the knowledge and now needs to recollect it. In learning the theorem, we are, in effect, recalling what we already know.

Plato famously illustrates the doctrine with an exchange between Socrates and a young slave, in which Socrates guides the slave from ignorance to mathematical knowledge. The slave's experiences, in the form of Socrates' questions and illustrations, are the occasion for his recollection of what he learned previously. Plato's metaphysics provides additional support for the Innate Knowledge Thesis. Since our knowledge is of abstract, eternal Forms which clearly lie beyond our sensory experience, it is *a priori*.

Contemporary supporters of Plato's position are scarce. The initial paradox, which Plato describes as a "trick argument" (*Meno*, 80e), rings sophistical. The metaphysical assumptions in the solution need justification. The solution does not answer the basic question: Just how did the slave's soul learn the theorem? The Intuition/Deduction thesis offers an equally, if not more, plausible account of how the slave gains knowledge *a priori*. Nonetheless, Plato's position illustrates the kind of reasoning that has caused many philosophers to adopt some form of the Innate Knowledge thesis. We are confident that we know certain propositions about the external world, but there seems to be no adequate explanation of how we gained this knowledge short of saying that it is innate. Its content is beyond what we directly gain in experience, as well

as what we can gain by performing mental operations on what experience provides. It does not seem to be based on an intuition or deduction. That it is innate in us appears to be the best explanation.

Noam Chomsky argues along similar lines in presenting what he describes as a "rationalist conception of the nature of language" (1975, p. 129). Chomsky argues that the experiences available to language learners are far too sparse to account for their knowledge of their language. To explain language acquisition, we must assume that learners have an innate knowledge of a universal grammar capturing the common deep structure of natural languages. It is important to note that Chomsky's language learners do not know particular propositions describing a universal grammar. They have a set of innate capacities or dispositions which enable and determine their language development. Chomsky gives us a theory of innate learning capacities or structures rather than a theory of innate knowledge. His view does not support the Innate Knowledge thesis as rationalists have traditionally understood it. As one commentator puts it, "Chomsky's principles … are innate neither in the sense that we are explicitly aware of them, nor in the sense that we have a disposition to recognize their truth as obvious under appropriate circumstances. And hence it is by no means clear that Chomsky is correct in seeing his theory as following the traditional rationalist account of the acquisition of knowledge" (Cottingham 1984, p. 124).

Peter Carruthers (1992) argues that we have innate knowledge of the principles of folk-psychology. Folk-psychology is a network of common-sense generalizations that hold independently of context or culture and concern the relationships of mental states to one another, to the environment and states of the body and to behavior (1992, p. 115). It includes such beliefs as that pains tend to be caused by injury, that pains tend to prevent us from concentrating on tasks, and that perceptions are generally caused by the appropriate state of the environment. Carruthers

notes the complexity of folk-psychology, along with its success in explaining our behavior and the fact that its explanations appeal to such unobservables as beliefs, desires, feelings and thoughts. He argues that the complexity, universality and depth of folk-psychological principles outstrips what experience can provide, especially to young children who by their fifth year already know a great many of them. This knowledge is also not the result of intuition or deduction; folk-psychological generalizations are not seen to be true in an act of intellectual insight. Carruthers concludes, "[The problem] concerning the child's acquisition of psychological generalizations cannot be solved, unless we suppose that much of folk-psychology is already innate, triggered locally by the child's experience of itself and others, rather than learned" (1992, p. 121).

Empiricists, and some rationalists, attack the Innate Knowledge thesis in two main ways. First, they offer accounts of how sense experience or intuition and deduction provide the knowledge that is claimed to be innate. Second, they directly criticize the Innate Knowledge thesis itself. The classic statement of this second line of attack is presented in Locke 1690. Locke raises the issue of just what innate knowledge is. Particular instances of knowledge are supposed to be in our minds as part of our rational make-up, but how are they "in our minds"? If the implication is that we all consciously have this knowledge, it is plainly false. Propositions often given as examples of innate knowledge, even such plausible candidates as the principle that the same thing cannot both be and not be, are not consciously accepted by children and those with severe cognitive limitations. If the point of calling such principles "innate" is not to imply that they are or have been consciously accepted by all rational beings, then it is hard to see what the point is. "No proposition can be said to be in the mind, which it never yet knew, which it never yet was conscious of" (1690, Book I, Chapter II, Section 5, p. 61). Proponents of innate knowledge might respond that some knowledge is innate in that we have the capacity to have it. That claim, while true, is of little interest,

however. "If the capacity of knowing, be the natural impression contended for, all the truths a man ever comes to know, will, by this account, be every one of them, innate; and this great point will amount to no more, but only an improper way of speaking; which whilst it pretends to assert the contrary, says nothing different from those, who deny innate principles. For nobody, I think, ever denied, that the mind was capable of knowing several truths" (1690, Book I, Chapter II, Section 5, p. 61). Locke thus challenges defenders of the Innate Knowledge thesis to present an account of innate knowledge that allows their position to be both true and interesting. A narrow interpretation of innateness faces counterexamples of rational individuals who do not meet its conditions. A generous interpretation implies that all our knowledge, even that clearly provided by experience, is innate.

Defenders of innate knowledge take up Locke's challenge. Leibniz responds (1704) by appealing to an account of innateness in terms of natural potential to avoid Locke's dilemma. Consider Peter Carruthers' similar reply.

> We have noted that while one form of nativism claims (somewhat implausibly) that knowledge is innate in the sense of being present as such (or at least in propositional form) from birth, it might also be maintained that knowledge is innate in the sense of being innately determined to make its appearance at some stage in childhood. This latter thesis is surely the most plausible version of nativism. (1992, p. 51)

Carruthers claims that our innate knowledge is determined through evolutionary selection (p. 111). Evolution has resulted in our being determined to know certain things (e.g. principles of folk-psychology) at particular stages of our life, as part of our natural development. Experiences provide the occasion for our consciously believing the known

propositions but not the basis for our knowledge of them (p. 52). Carruthers thus has a ready reply to Locke's counterexamples of children and cognitively limited persons who do not believe propositions claimed to be instances of innate knowledge. The former have not yet reached the proper stage of development; the latter are persons in whom natural development has broken down (pp. 49–50).

A serious problem for the Innate Knowledge thesis remains, however. We know a proposition only if it is true, we believe it and our belief is warranted. Rationalists who assert the existence of innate knowledge are not just claiming that, as a matter of human evolution, God's design or some other factor, at a particular point in our development, certain sorts of experiences trigger our belief in particular propositions in a way that does not involve our learning them from the experiences. Their claim is even bolder: In at least some of these cases, our empirically triggered, but not empirically warranted, belief is nonetheless warranted and so known. How can these beliefs be warranted if they do not gain their warrant from the experiences that cause us to have them or from intuition and deduction?

Some rationalists think that a reliabilist account of warrant provides the answer. According to Reliabilism, beliefs are warranted if they are formed by a process that generally produces true beliefs rather than false ones. The true beliefs that constitute our innate knowledge are warranted, then, because they are formed as the result of a reliable belief-forming process. Carruthers maintains that "Innate beliefs will count as known provided that the process through which they come to be innate is a reliable one (provided, that is, that the process tends to generate beliefs that are true)" (1992, p. 77). He argues that natural selection results in the formation of some beliefs and is a truth-reliable process.

An appeal to Reliabilism, or a similar causal theory of warrant, may well be the best way for rationalists to develop the Innate Knowledge thesis.

They have a difficult row to hoe, however. First, such accounts of warrant are themselves quite controversial. Second, rationalists must give an account of innate knowledge that maintains and explains the distinction between innate knowledge and *a posteriori* knowledge, and it is not clear that they will be able to do so within such an account of warrant. Suppose for the sake of argument that we have innate knowledge of some proposition, *P*. What makes our knowledge that *P* innate? To sharpen the question, what difference between our knowledge that *P* and a clear case of *a posteriori* knowledge, say our knowledge that something is red based on our current visual experience of a red table, makes the former innate and the latter not innate? In each case, we have a true, warranted belief. In each case, presumably, our belief gains its warrant from the fact that it meets a particular causal condition, e.g., it is produced by a reliable process. In each case, the causal process is one in which an experience causes us to believe the proposition at hand (that *P*; that something is red), for, as defenders of innate knowledge admit, our belief that *P* is "triggered" by an experience, as is our belief that something is red. The insight behind the Innate Knowledge thesis seems to be that the difference between our innate and *a posteriori* knowledge lies in the relation between our experience and our belief in each case. The experience that causes our belief that *P* does not "contain" the information that *P*, while our visual experience of a red table does "contain" the information that something is red. Yet, exactly what is the nature of this containment relation between our experiences, on the one hand, and what we believe, on the other, that is missing in the one case but present in the other? The nature of the experience-belief relation seems quite similar in each. The causal relation between the experience that triggers our belief that *P* and our belief that *P* is contingent, as is the fact that the belief-forming process is reliable. The same is true of our experience of a red table and our belief that something is red. The causal relation between the experience and our belief is again contingent. We might have been so constructed that the experience we

describe as "being appeared to redly" caused us to believe, not that something is red, but that something is hot. The process that takes us from the experince to our belief is also only contingently reliable. Moreover, if our experience of a red table "contains" the information that something is red, then that fact, not the existence of a reliable belief-forming process between the two, should be the reason why the experience warrants our belief. By appealing to Reliablism, or some other causal theory of warrant, rationalists may obtain a way to explain how innate knowledge can be warranted. They still need to show how their explanation supports an account of the difference between innate knowledge and *a posteriori* knowledge.

4. The Innate Concept Thesis

According to the Innate Concept thesis, some of our concepts have not been gained from experience. They are instead part of our rational make-up, and experience simply triggers a process by which we consciously grasp them. The main concern motivating the rationalist should be familiar by now: the content of some concepts seems to outstrip anything we could have gained from experience. An example of this reasoning is presented by Descartes in the *Meditations*. Although he sometimes seems committed to the view that all our ideas are innate (Adams 1975 and Gotham 2002), he there classifies our ideas as adventitious, invented by us, and innate. Adventitious ideas, such as a sensation of heat, are gained directly through sense experience. Ideas invented by us, such as our idea of a hippogriff, are created by us from other ideas we possess. Innate ideas, such as our ideas of God, of extended matter, of substance and of a perfect triangle, are placed in our minds by God at creation. Consider Descartes's argument that our concept of God, as an infinitely perfect being, is innate. Our concept of God is not directly gained in experience, as particular tastes, sensations and mental images might be. Its content is beyond what we

could ever construct by applying available mental operations to what experience directly provides. From experience, we can gain the concept of a being with finite amounts of various perfections, one, for example, that is finitely knowledgeable, powerful and good. We cannot however move from these empirical concepts to the concept of a being of infinite perfection. ("I must not think that, just as my conceptions of rest and darkness are arrived at by negating movement and light, so my perception of the infinite is arrived at not by means of a true idea but by merely negating the finite," Third Meditation, p. 94.) Descartes supplements this argument by another. Not only is the content of our concept of God beyond what experience can provide, the concept is a prerequisite for our employment of the concept of finite perfection gained from experience. ("My perception of the infinite, that is God, is in some way prior to my perception of the finite, that is myself. For how could I understand that I doubted or desired—that is lacked something—and that I was not wholly perfect, unless there were in me some idea of a more perfect being which enabled me to recognize my own defects by comparison," Third Meditation, p. 94).

An empiricist response to this general line of argument is given by Locke (1690, Book I, Chapter IV, Sections 1–25, pp. 91–107). First, there is the problem of explaining what it is for someone to have an innate concept. If having an innate concept entails consciously entertaining it at present or in the past, then Descartes's position is open to obvious counterexamples. Young children and people from other cultures do not consciously entertain the concept of God and have not done so. Second, there is the objection that we have no need to appeal to innate concepts in the first place. Contrary to Descartes' argument, we can explain how experience provides all our ideas, including those the rationalists take to be innate, and with just the content that the rationalists attribute to them.

Leibniz (1704) offers a rationalist reply to the first concern. Where Locke puts forth the image of the mind as a blank tablet on which experience writes, Leibniz offers us the image of a block of marble, the veins of which determine what sculpted figures it will accept.

> This is why I have taken as an illustration a block of veined marble, rather than a wholly uniform block or blank tablets, that is to say what is called tabula rasa in the language of the philosophers. For if the soul were like those blank tablets, truths would be in us in the same way as the figure of Hercules is in a block of marble, when the marble is completely indifferent whether it receives this or some other figure. But if there were veins in the stone which marked out the figure of Hercules rather than other figures, this stone would be more determined thereto, and Hercules would be as it were in some manner innate in it, although labour would be needed to uncover the veins, and to clear them by polishing, and by cutting away what prevents them from appearing. It is in this way that ideas and truths are innate in us, like natural inclinations and dispositions, natural habits or potentialities, and not like activities, although these potentialities are always accompanied by some activities which correspond to them, though they are often imperceptible. (1704, Preface, p. 153)

Leibniz's metaphor contains an insight that Locke misses. The mind plays a role in determining the nature of its contents. This point does not, however, require the adoption of the Innate Concept thesis.

Rationalists have responded to the second part of the empiricist attack on the Innate Concept thesis—the empricists' claim that the thesis is without basis, as all our ideas can be explained as derived from experience—by focusing on difficulties in the empiricists' attempts to give such an explanation. The difficulties are illustrated by Locke's account. According

to Locke, experience consists in external sensation and inner reflection. All our ideas are either simple or complex, with the former being received by us passively in sensation or reflection and the latter being built by the mind from simple materials through various mental operations. Right at the start, the account of how simple ideas are gained is open to an obvious counterexample acknowledged, but then set aside, by Hume in presenting his own empiricist theory. Consider the mental image of a particular shade of blue. If Locke is right, the idea is a simple one and should be passively received by the mind through experience. Hume points out otherwise.

> Suppose therefore a person to have enjoyed his sight for thirty years and to have become perfectly acquainted with colors of all kinds, except one particular shade of blue, for instance, which it never has been his fortune to meet with; let all the different shades of that color, except that single one, be placed before him, descending gradually from the deepest to the lightest, it is plain that he will perceive a blank where that shade is wanting and will be sensible that there is a greater distance in that place between the contiguous colors than in any other. Now I ask whether it be possible for him, from his own imagination, to supply this deficiency and raise up to himself the idea of that particular shade, though it had never been conveyed to him by his senses? I believe there are but few will be of the opinion that he can… (1748, Section II, pp. 29–30)

Even when it comes to such simple ideas as the image of a particular shade of blue, the mind is more than a blank slate on which experience writes.

Consider too our concept of a particular color, say red. Critics of Locke's account have pointed out the weaknesses in his explanation of how we gain such a concept by the mental operation of abstraction on individual

cases. For one thing, it makes the incorrect assumption that various instances of a particular concept share a common feature. Carruthers puts the objection as follows.

> In fact problems arise for empiricists even in connection with the very simplest concepts, such as those of colour. For it is false that all instances of a given colour share some common feature. In which case we cannot acquire the concept of that colour by abstracting the common feature of our experience. Thus consider the concept *red*. Do all shades of red have something in common? If so, what? It is surely false that individual shades of red consist, as it were, of two distinguishable elements a general redness together with a particular shade. Rather, redness consists in a continuous *range* of shades, each of which is only just distinguishable from its neighbors. Acquiring the concept *red* is a matter of learning the extent of the range. (1992, p. 59)

For another thing, Locke's account of concept acquisition from particular experiences seems circular.

> As it stands, however, Locke's account of concept acquisition appears viciously circular. For noticing or attending to a common feature of various things presupposes that you already possess the concept of the feature in question. (Carruthers 1992, p. 55)

Consider in this regard Locke's account of how we gain our concept of causation.

> In the notice that our senses take of the constant vicissitude of things, we cannot but observe, that several particulars, both qualities and substances; begin to exist; and that they receive this their existence from the due application and operation of some other being. From this observation, we get our ideas of cause and

effect. (1690, Book II, Chapter 26, Section 1, pp. 292–293)

We get our concept of causation from our observation that some things receive their existence from the application and operation of some other things. Yet, we cannot make this observation unless we already have the concept of causation. Locke's account of how we gain our idea of power displays a similar circularity.

> The mind being every day informed, by the senses, of the alteration of those simple ideas, it observes in things without; and taking notice how one comes to an end, and ceases to be, and another begins to exist which was not before; reflecting also on what passes within itself, and observing a constant change of its ideas, sometimes by the impression of outward objects on the senses, and sometimes by the determination of its own choice; and concluding from what it has so constantly observed to have been, that the like changes will for the future be made in the same things, by like agents, and by the like ways, considers in one thing the possibility of having any of its simple ideas changed, and in another the possibility of making that change; and so comes by that idea which we call power. (1690, Chapter XXI, Section 1, pp. 219–220)

We come by the idea of power though considering the possibility of changes in our ideas made by experiences and our own choices. Yet, to consider this possibility—of some things *making* a change in others—we must already have a concept of power.

One way to meet at least some of these challenges to an empiricist account of the origin of our concepts is to revise our understanding of the content of our concepts so as to bring them more in line with what experience will clearly provide. Hume famously takes this approach. Beginning in a way

reminiscent of Locke, he distinguishes between two forms of mental contents or "perceptions," as he calls them: impressions and ideas. Impressions are the contents of our current experiences: our sensations, feelings, emotions, desires, and so on. Ideas are mental contents derived from impressions. Simple ideas are copies of impressions; complex ideas are derived from impressions by "compounding, transposing, augmenting or diminishing" them. Given that all our ideas are thus gained from experience, Hume offers us the following method for determining the content of any idea and thereby the meaning of any term taken to express it.

> When we entertain, therefore, any suspicion that a philosophical term is employed without any meaning or idea (as is but too frequent), we need but inquire *from what impression is that supposed idea derived*? And if it be impossible to assign any, this will confirm our suspicion. (1748, Section II, p. 30)

Using this test, Hume draws out one of the most important implications of the empiricists' denial of the Innate Concept thesis. If experience is indeed the source of all ideas, then our experiences also determine the content of our ideas. Our ideas of causation, of substance, of right and wrong have their content determined by the experiences that provide them. Those experiences, Hume argues, are unable to support the content that many rationalists and some empiricists, such as Locke, attribute to the corresponding ideas. Our inability to explain how some concepts, with the contents the rationalists attribute to them, are gained from experience should not lead us to adopt the Innate Concept thesis. It should lead us to accept a more limited view of the contents for those concepts, and thereby a more limited view of our ability to describe and understand the world.

Consider, for example, our idea of causation. Descartes takes it to be innate. Locke offers an apparently circular account of how it is gained

from experience. Hume's empiricist account severely limits its content. Our idea of causation is derived from a feeling of expectation rooted in our experiences of the constant conjunction of similar causes and effects.

> It appears, then, that this idea of a necessary connection among events arises from a number of similar instances which occur, of the constant conjunction of these events; nor can that idea ever be suggested by any one of these instances surveyed in all possible lights and positions. But there is nothing in a number of instances, different from every single instance, which is supposed to be exactly similar, except only that after a repetition of similar instances the mind is carried by habit, upon the appearance of one event, to expect its usual attendant and to believe that it will exist. This connection, therefore, which we *feel* in the mind, this customary transition of the imagination from one object to its usual attendant, is the sentiment or impression from which we form the idea of power or necessary connection. (1748, Section VII, Part 2, p. 86)

The source of our idea in experience determines its content.

> Suitably to this experience, therefore, we may define a cause to be an object followed by another, and where all the objects, similar to the first are followed by objects similar to the second... We may, therefore, suitably to this experience, form another definition of cause and call it *an object followed by another, and whose appearance always conveys the thought of the other*. (1748, Section VII, Part 2, p. 87)

Our claims, and any knowledge we may have, about causal connections in the world turn out, given the limited content of our empirically based concept of causation, to be claims and knowledge about the constant

conjunction of events and our own feelings of expectation. Thus, the initial disagreement between rationalists and empiricists about the source of our ideas leads to one about their content and thereby the content of our descriptions and knowledge of the world.

Like philosophical debates generally, the rationalist/empiricist debate ultimately concerns our position in the world, in this case our position as rational inquirers. To what extent do our faculties of reason and experience support our attempts to know and understand our situation?

Bibliography

Works Cited

Adams, R., 1975, "Where Do Our Ideas Come From? Descartes vs Locke", reprinted in Stitch S. (ed.) *Innate Ideas*, Berkeley, CA: California University Press.

Alexander, J. and Weinberg, J., 2007, "Analytic Epistemology and Experimental Philosophy," *Philosophy Compass*, 2(1): 56–80.

Ayer, A. J., 1952, *Language, Truth and Logic*, New York: Dover Publications.

Bealer, G., 1999, "A Theory of the *A priori*," *Noûs*, 33: 29–55.

Carruthers, P., 1992, *Human Knowledge and Human Nature*, Oxford: Oxford University Press.

Descartes, R., 1628, *Rules for the Direction of our Native Intelligence*, in *Descartes: Selected Philosophical Writings*, transl. John Cottingham, Robert Stoothoff and Dugald Murdoch, Cambridge: Cambridge University Press, 1988.

—, 1641, *Meditations*, in *Descartes: Selected Philosophical Writings*, transl. John Cottingham, Robert Stoothoff and Dugald Murdoch, Cambridge: Cambridge University Press, 1988.

—, 1644, *Principles of Philosophy*, in *Descartes: Selected Philosophical*

Immanuel Kant

First published Thu May 20, 2010; substantive revision Mon Jan 25, 2016

Immanuel Kant (1724–1804) is the central figure in modern philosophy. He synthesized early modern rationalism and empiricism, set the terms for much of nineteenth and twentieth century philosophy, and continues to exercise a significant influence today in metaphysics, epistemology, ethics, political philosophy, aesthetics, and other fields. The fundamental idea of Kant's "critical philosophy" – especially in his three Critiques: the Critique of Pure Reason (1781, 1787), the Critique of Practical Reason (1788), and the Critique of the Power of Judgment (1790) – is human autonomy. He argues that the human understanding is the source of the general laws of nature that structure all our experience; and that human reason gives itself the moral law, which is our basis for belief in God, freedom, and immortality. Therefore, scientific knowledge, morality, and religious belief are mutually consistent and secure because they all rest on the same foundation of human autonomy, which is also the final end of nature according to the teleological worldview of reflecting judgment that Kant introduces to unify the theoretical and practical parts of his philosophical system.

- 1. Life and works
- 2. Kant's project in the Critique of Pure Reason
    - 2.1 The crisis of the Enlightenment
    - 2.2 Kant's Copernican revolution in philosophy
- 3. Transcendental idealism
    - 3.1 The two-objects interpretation
    - 3.2 The two-aspects interpretation
- 4. The transcendental deduction
    - 4.1 Self-consciousness
    - 4.2 Objectivity and judgment

1. Life and works

Immanuel Kant was born April 22, 1724 in Königsberg, near the southeastern shore of the Baltic Sea. Today Königsberg has been renamed Kaliningrad and is part of Russia. But during Kant's lifetime Königsberg was the capital of East Prussia, and its dominant language was German. Though geographically remote from the rest of Prussia and other German cities, Königsberg was then a major commercial center, an important military port, and a relatively cosmopolitan university town.[1]

Kant was born into an artisan family of modest means. His father was a master harness maker, and his mother was the daughter of a harness

maker, though she was better educated than most women of her social class. Kant's family was never destitute, but his father's trade was in decline during Kant's youth and his parents at times had to rely on extended family for financial support.

Kant's parents were Pietist and he attended a Pietist school, the Collegium Fridericianum, from ages eight through fifteen. Pietism was an evangelical Lutheran movement that emphasized conversion, reliance on divine grace, the experience of religious emotions, and personal devotion involving regular Bible study, prayer, and introspection. Kant reacted strongly against the forced soul-searching to which he was subjected at the Collegium Fridericianum, in response to which he sought refuge in the Latin classics, which were central to the school's curriculum. Later the mature Kant's emphasis on reason and autonomy, rather than emotion and dependence on either authority or grace, may in part reflect his youthful reaction against Pietism. But although the young Kant loathed his Pietist schooling, he had deep respect and admiration for his parents, especially his mother, whose "genuine religiosity" he described as "not at all enthusiastic." According to his biographer, Manfred Kuehn, Kant's parents probably influenced him much less through their Pietism than through their artisan values of "hard work, honesty, cleanliness, and independence," which they taught him by example.[2]

Kant attended college at the University of Königsberg, known as the Albertina, where his early interest in classics was quickly superseded by philosophy, which all first year students studied and which encompassed mathematics and physics as well as logic, metaphysics, ethics, and natural law. Kant's philosophy professors exposed him to the approach of Christian Wolff (1679–1750), whose critical synthesis of the philosophy of G. W. Leibniz (1646–1716) was then very influential in German universities. But Kant was also exposed to a range of German and British critics of Wolff, and there were strong doses of Aristotelianism and

Pietism represented in the philosophy faculty as well. Kant's favorite teacher was Martin Knutzen (1713–1751), a Pietist who was heavily influenced by both Wolff and the English philosopher John Locke (1632–1704). Knutzen introduced Kant to the work of Isaac Newton (1642–1727), and his influence is visible in Kant's first published work, Thoughts on the True Estimation of Living Forces (1747), which was a critical attempt to mediate a dispute in natural philosophy between Leibnizians and Newtonians over the proper measurement of force.

After college Kant spent six years as a private tutor to young children outside Königsberg. By this time both of his parents had died and Kant's finances were not yet secure enough for him to pursue an academic career. He finally returned to Königsberg in 1754 and began teaching at the Albertina the following year. For the next four decades Kant taught philosophy there, until his retirement from teaching in 1796 at the age of seventy-two.

Kant had a burst of publishing activity in the years after he returned from working as a private tutor. In 1754 and 1755 he published three scientific works – one of which, Universal Natural History and Theory of the Heavens (1755), was a major book in which, among other things, he developed what later became known as the nebular hypothesis about the formation of the solar system. Unfortunately, the printer went bankrupt and the book had little immediate impact. To secure qualifications for teaching at the university, Kant also wrote two Latin dissertations: the first, entitled Concise Outline of Some Reflections on Fire (1755), earned him the Magister degree; and the second, New Elucidation of the First Principles of Metaphysical Cognition (1755), entitled him to teach as an unsalaried lecturer. The following year he published another Latin work, The Employment in Natural Philosophy of Metaphysics Combined with Geometry, of Which Sample I Contains the Physical Monadology (1756), in hopes of succeeding Knutzen as associate professor of logic and

metaphysics, though Kant failed to secure this position. Both the New Elucidation, which was Kant's first work concerned mainly with metaphysics, and the Physical Monadology further develop the position on the interaction of finite substances that he first outlined in Living Forces. Both works depart from Leibniz-Wolffian views, though not radically. The New Elucidation in particular shows the influence of Christian August Crusius (1715–1775), a German critic of Wolff.[3]

As an unsalaried lecturer at the Albertina Kant was paid directly by the students who attended his lectures, so he needed to teach an enormous amount and to attract many students in order to earn a living. Kant held this position from 1755 to 1770, during which period he would lecture an average of twenty hours per week on logic, metaphysics, and ethics, as well as mathematics, physics, and physical geography. In his lectures Kant used textbooks by Wolffian authors such as Alexander Gottlieb Baumgarten (1714–1762) and Georg Friedrich Meier (1718–1777), but he followed them loosely and used them to structure his own reflections, which drew on a wide range of ideas of contemporary interest. These ideas often stemmed from British sentimentalist philosophers such as David Hume (1711–1776) and Francis Hutcheson (1694–1747), some of whose texts were translated into German in the mid-1750s; and from the Swiss philosopher Jean-Jacques Rousseau (1712–1778), who published a flurry of works in the early 1760s. From early in his career Kant was a popular and successful lecturer. He also quickly developed a local reputation as a promising young intellectual and cut a dashing figure in Königsberg society.

After several years of relative quiet, Kant unleashed another burst of publications in 1762–1764, including five philosophical works. The False Subtlety of the Four Syllogistic Figures (1762) rehearses criticisms of Aristotelian logic that were developed by other German philosophers. The Only Possible Argument in Support of a Demonstration of the Existence

of God (1762–3) is a major book in which Kant drew on his earlier work in Universal History and New Elucidation to develop an original argument for God's existence as a condition of the internal possibility of all things, while criticizing other arguments for God's existence. The book attracted several positive and some negative reviews. In 1762 Kant also submitted an essay entitled Inquiry Concerning the Distinctness of the Principles of Natural Theology and Morality to a prize competition by the Prussian Royal Academy, though Kant's submission took second prize to Moses Mendelssohn's winning essay (and was published with it in 1764). Kant's Prize Essay, as it is known, departs more significantly from Leibniz-Wolffian views than his earlier work and also contains his first extended discussion of moral philosophy in print. The Prize Essay draws on British sources to criticize German rationalism in two respects: first, drawing on Newton, Kant distinguishes between the methods of mathematics and philosophy; and second, drawing on Hutcheson, he claims that "an unanalysable feeling of the good" supplies the material content of our moral obligations, which cannot be demonstrated in a purely intellectual way from the formal principle of perfection alone (2:299).[4] These themes reappear in the Attempt to Introduce the Concept of Negative Magnitudes into Philosophy (1763), whose main thesis, however, is that the real opposition of conflicting forces, as in causal relations, is not reducible to the logical relation of contradiction, as Leibnizians held. In Negative Magnitudes Kant also argues that the morality of an action is a function of the internal forces that motivate one to act, rather than of the external (physical) actions or their consequences. Finally, Observations on the Feeling of the Beautiful and the Sublime (1764) deals mainly with alleged differences in the tastes of men and women and of people from different cultures. After it was published, Kant filled his own interleaved copy of this book with (often unrelated) handwritten remarks, many of which reflect the deep influence of Rousseau on his thinking about moral philosophy in the mid-1760s.

These works helped to secure Kant a broader reputation in Germany, but for the most part they were not strikingly original. Like other German philosophers at the time, Kant's early works are generally concerned with using insights from British empiricist authors to reform or broaden the German rationalist tradition without radically undermining its foundations. While some of his early works tend to emphasize rationalist ideas, others have a more empiricist emphasis. During this time Kant was striving to work out an independent position, but before the 1770s his views remained fluid.

In 1766 Kant published his first work concerned with the possibility of metaphysics, which later became a central topic of his mature philosophy. Dreams of a Spirit-Seer Elucidated by Dreams of Metaphysics, which he wrote soon after publishing a short Essay on Maladies of the Mind (1764), was occasioned by Kant's fascination with the Swedish visionary Emanuel Swedenborg (1688–1772), who claimed to have insight into a spirit world that enabled him to make a series of apparently miraculous predictions. In this curious work Kant satirically compares Swedenborg's spirit-visions to the belief of rationalist metaphysicians in an immaterial soul that survives death, and he concludes that philosophical knowledge of either is impossible because human reason is limited to experience. The skeptical tone of Dreams is tempered, however, by Kant's suggestion that "moral faith" nevertheless supports belief in an immaterial and immortal soul, even if it is not possible to attain metaphysical knowledge in this domain (2:373).

In 1770, at the age of forty-six, Kant was appointed to the chair in logic and metaphysics at the Albertina, after teaching for fifteen years as an unsalaried lecturer and working since 1766 as a sublibrarian to supplement his income. Kant was turned down for the same position in 1758. But later, as his reputation grew, he declined chairs in philosophy at Erlangen (1769) and Jena (1770) in hopes of obtaining one in Königsberg. After

Kant was finally promoted, he gradually extended his repertoire of lectures to include anthropology (Kant's was the first such course in Germany and became very popular), rational theology, pedagogy, natural right, and even mineralogy and military fortifications. In order to inaugurate his new position, Kant also wrote one more Latin dissertation: Concerning the Form and Principles of the Sensible and Intelligible World (1770), which is known as the Inaugural Dissertation.

The Inaugural Dissertation departs more radically from both Wolffian rationalism and British sentimentalism than Kant's earlier work. Inspired by Crusius and the Swiss natural philosopher Johann Heinrich Lambert (1728–1777), Kant distinguishes between two fundamental powers of cognition, sensibility and understanding (intelligence), where the Leibniz-Wolffians regarded understanding (intellect) as the only fundamental power. Kant therefore rejects the rationalist view that sensibility is only a confused species of intellectual cognition, and he replaces this with his own view that sensibility is distinct from understanding and brings to perception its own subjective forms of space and time – a view that developed out of Kant's earlier criticism of Leibniz's relational view of space in Concerning the Ultimate Ground of the Differentiation of Directions in Space (1768). Moreover, as the title of the Inaugural Dissertation indicates, Kant argues that sensibility and understanding are directed at two different worlds: sensibility gives us access to the sensible world, while understanding enables us to grasp a distinct intelligible world. These two worlds are related in that what the understanding grasps in the intelligible world is the "paradigm" of "NOUMENAL PERFECTION," which is "a common measure for all other things in so far as they are realities." Considered theoretically, this intelligible paradigm of perfection is God; considered practically, it is "MORAL PERFECTION" (2:396). The Inaugural Dissertation thus develops a form of Platonism; and it rejects the view of British sentimentalists that moral

judgments are based on feelings of pleasure or pain, since Kant now holds that moral judgments are based on pure understanding alone.

After 1770 Kant never surrendered the views that sensibility and understanding are distinct powers of cognition, that space and time are subjective forms of human sensibility, and that moral judgments are based on pure understanding (or reason) alone. But his embrace of Platonism in the Inaugural Dissertation was short-lived. He soon denied that our understanding is capable of insight into an intelligible world, which cleared the path toward his mature position in the Critique of Pure Reason (1781), according to which the understanding (like sensibility) supplies forms that structure our experience of the sensible world, to which human knowledge is limited, while the intelligible (or noumenal) world is strictly unknowable to us. Kant spent a decade working on the Critique of Pure Reason and published nothing else of significance between 1770 and 1781. But its publication marked the beginning of another burst of activity that produced Kant's most important and enduring works. Because early reviews of the Critique of Pure Reason were few and (in Kant's judgment) uncomprehending, he tried to clarify its main points in the much shorter Prolegomena to Any Future Metaphysics That Will Be Able to Come Forward as a Science (1783). Among the major books that rapidly followed are the Groundwork of the Metaphysics of Morals (1785), Kant's main work on the fundamental principle of morality; the Metaphysical Foundations of Natural Science (1786), his main work on natural philosophy in what scholars call his critical period (1781–1798); the second and substantially revised edition of the Critique of Pure Reason (1787); the Critique of Practical Reason (1788), a fuller discussion of topics in moral philosophy that builds on (and in some ways revises) the Groundwork; and the Critique of the Power of Judgment (1790), which deals with aesthetics and teleology. Kant also published a number of important essays in this period, including Idea for a Universal History With a Cosmopolitan Aim (1784) and Conjectural Beginning of Human

History (1786), his main contributions to the philosophy of history; An Answer to the Question: What is Enlightenment? (1784), which broaches some of the key ideas of his later political essays; and What Does it Mean to Orient Oneself in Thinking? (1786), Kant's intervention in the pantheism controversy that raged in German intellectual circles after F. H. Jacobi (1743–1819) accused the recently deceased G. E. Lessing (1729–1781) of Spinozism.

With these works Kant secured international fame and came to dominate German philosophy in the late 1780s. But in 1790 he announced that the Critique of the Power of Judgment brought his critical enterprise to an end (5:170). By then K. L. Reinhold (1758–1823), whose Letters on the Kantian Philosophy (1786) popularized Kant's moral and religious ideas, had been installed (in 1787) in a chair devoted to Kantian philosophy at Jena, which was more centrally located than Königsberg and rapidly developing into the focal point of the next phase in German intellectual history. Reinhold soon began to criticize and move away from Kant's views. In 1794 his chair at Jena passed to J. G. Fichte, who had visited the master in Königsberg and whose first book, Attempt at a Critique of All Revelation (1792), was published anonymously and initially mistaken for a work by Kant himself. This catapulted Fichte to fame, but he too soon moved away from Kant and developed an original position quite at odds with Kant's, which Kant finally repudiated publicly in 1799 (12:370–371). Yet while German philosophy moved on to assess and respond to Kant's legacy, Kant himself continued publishing important works in the 1790s. Among these are Religion Within the Boundaries of Mere Reason (1793), which drew a censure from the Prussian King when Kant published the book after its second essay was rejected by the censor; The Conflict of the Faculties (1798), a collection of essays inspired by Kant's troubles with the censor and dealing with the relationship between the philosophical and theological faculties of the university; On the Common Saying: That May be Correct in Theory, But it is of No Use in Practice (1793), Toward

Perpetual Peace (1795), and the Doctrine of Right, the first part of the Metaphysics of Morals (1797), Kant's main works in political philosophy; the Doctrine of Virtue, the second part of the Metaphysics of Morals (1797), a catalogue of duties that Kant had been planning for more than thirty years; and Anthropology From a Pragmatic Point of View (1798), based on Kant's anthropology lectures. Several other compilations of Kant's lecture notes from other courses were published later, but these were not prepared by Kant himself.

Kant retired from teaching in 1796. For nearly two decades he had lived a highly disciplined life focused primarily on completing his philosophical system, which began to take definite shape in his mind only in middle age. After retiring he came to believe that there was a gap in this system separating the metaphysical foundations of natural science from physics itself, and he set out to close this gap in a series of notes that postulate the existence of an ether or caloric matter. These notes, known as the Opus Postumum, remained unfinished and unpublished in Kant's lifetime, and scholars disagree on their significance and relation to his earlier work. It is clear, however, that these late notes show unmistakable signs of Kant's mental decline, which became tragically precipitous around 1800. Kant died February 12, 1804, just short of his eightieth birthday.

2. Kant's project in the Critique of Pure Reason

The main topic of the Critique of Pure Reason is the possibility of metaphysics, understood in a specific way. Kant defines metaphysics in terms of "the cognitions after which reason might strive independently of all experience," and his goal in the book is to reach a "decision about the possibility or impossibility of a metaphysics in general, and the determination of its sources, as well as its extent and boundaries, all, however, from principles" (Axii. See also Bxiv; and 4:255–257). Thus metaphysics for Kant concerns a priori knowledge, or knowledge whose

justification does not depend on experience; and he associates a priori knowledge with reason. The project of the Critique is to examine whether, how, and to what extent human reason is capable of a priori knowledge.

2.1 The crisis of the Enlightenment

To understand the project of the Critique better, let us consider the historical and intellectual context in which it was written.[5] Kant wrote the Critique toward the end of the Enlightenment, which was then in a state of crisis. Hindsight enables us to see that the 1780's was a transitional decade in which the cultural balance shifted decisively away from the Enlightenment toward Romanticism, but of course Kant did not have the benefit of such hindsight.

The Enlightenment was a reaction to the rise and successes of modern science in the sixteenth and seventeenth centuries. The spectacular achievement of Newton in particular engendered widespread confidence and optimism about the power of human reason to control nature and to improve human life. One effect of this new confidence in reason was that traditional authorities were increasingly questioned. For why should we need political or religious authorities to tell us how to live or what to believe, if each of us has the capacity to figure these things out for ourselves? Kant expresses this Enlightenment commitment to the sovereignty of reason in the Critique:

> Our age is the age of criticism, to which everything must submit. Religion through its holiness and legislation through its majesty commonly seek to exempt themselves from it. But in this way they excite a just suspicion against themselves, and cannot lay claim to that unfeigned respect that reason grants only to that which has been able to withstand its free and public examination (Axi).

Enlightenment is about thinking for oneself rather than letting others think for you, according to What is Enlightenment? (8:35). In this essay, Kant also expresses the Enlightenment faith in the inevitability of progress. A few independent thinkers will gradually inspire a broader cultural movement, which ultimately will lead to greater freedom of action and governmental reform. A culture of enlightenment is "almost inevitable" if only there is "freedom to make public use of one's reason in all matters" (8:36).

The problem is that to some it seemed unclear whether progress would in fact ensue if reason enjoyed full sovereignty over traditional authorities; or whether unaided reasoning would instead lead straight to materialism, fatalism, atheism, skepticism (Bxxxiv), or even libertinism and authoritarianism (8:146). The Enlightenment commitment to the sovereignty of reason was tied to the expectation that it would not lead to any of these consequences but instead would support certain key beliefs that tradition had always sanctioned. Crucially, these included belief in God, the soul, freedom, and the compatibility of science with morality and religion. Although a few intellectuals rejected some or all of these beliefs, the general spirit of the Enlightenment was not so radical. The Enlightenment was about replacing traditional authorities with the authority of individual human reason, but it was not about overturning traditional moral and religious beliefs.

Yet the original inspiration for the Enlightenment was the new physics, which was mechanistic. If nature is entirely governed by mechanistic, causal laws, then it may seem that there is no room for freedom, a soul, or anything but matter in motion. This threatened the traditional view that morality requires freedom. We must be free in order to choose what is right over what is wrong, because otherwise we cannot be held responsible. It also threatened the traditional religious belief in a soul that can survive death or be resurrected in an afterlife. So modern science, the

pride of the Enlightenment, the source of its optimism about the powers of human reason, threatened to undermine traditional moral and religious beliefs that free rational thought was expected to support. This was the main intellectual crisis of the Enlightenment.

The Critique of Pure Reason is Kant's response to this crisis. Its main topic is metaphysics because, for Kant, metaphysics is the domain of reason – it is "the inventory of all we possess through pure reason, ordered systematically" (Axx) – and the authority of reason was in question. Kant's main goal is to show that a critique of reason by reason itself, unaided and unrestrained by traditional authorities, establishes a secure and consistent basis for both Newtonian science and traditional morality and religion. In other words, free rational inquiry adequately supports all of these essential human interests and shows them to be mutually consistent. So reason deserves the sovereignty attributed to it by the Enlightenment.

2.2 Kant's Copernican revolution in philosophy

To see how Kant attempts to achieve this goal in the Critique, it helps to reflect on his grounds for rejecting the Platonism of the Inaugural Dissertation. In a way the Inaugural Dissertation also tries to reconcile Newtonian science with traditional morality and religion, but its strategy is different from that of the Critique. According to the Inaugural Dissertation, Newtonian science is true of the sensible world, to which sensibility gives us access; and the understanding grasps principles of divine and moral perfection in a distinct intelligible world, which are paradigms for measuring everything in the sensible world. So on this view our knowledge of the intelligible world is a priori because it does not depend on sensibility, and this a priori knowledge furnishes principles for judging the sensible world because in some way the sensible world itself conforms to or imitates the intelligible world.

Soon after writing the Inaugural Dissertation, however, Kant expressed doubts about this view. As he explained in a February 21, 1772 letter to his friend and former student, Marcus Herz:

> In my dissertation I was content to explain the nature of intellectual representations in a merely negative way, namely, to state that they were not modifications of the soul brought about by the object. However, I silently passed over the further question of how a representation that refers to an object without being in any way affected by it can be possible.... [B]y what means are these [intellectual representations] given to us, if not by the way in which they affect us? And if such intellectual representations depend on our inner activity, whence comes the agreement that they are supposed to have with objects – objects that are nevertheless not possibly produced thereby?…[A]s to how my understanding may form for itself concepts of things completely a priori, with which concepts the things must necessarily agree, and as to how my understanding may formulate real principles concerning the possibility of such concepts, with which principles experience must be in exact agreement and which nevertheless are independent of experience – this question, of how the faculty of understanding achieves this conformity with the things themselves, is still left in a state of obscurity. (10:130–131)

Here Kant entertains doubts about how a priori knowledge of an intelligible world would be possible. The position of the Inaugural Dissertation is that the intelligible world is independent of the human understanding and of the sensible world, both of which (in different ways) conform to the intelligible world. But, leaving aside questions about what it means for the sensible world to conform to an intelligible world, how is it possible for the human understanding to conform to or grasp an intelligible world? If the intelligible world is independent of our

understanding, then it seems that we could grasp it only if we are passively affected by it in some way. But for Kant sensibility is our passive or receptive capacity to be affected by objects that are independent of us (2:392, A51/B75). So the only way we could grasp an intelligible world that is independent of us is through sensibility, which means that our knowledge of it could not be a priori. The pure understanding alone could at best enable us to form representations of an intelligible world. But since these intellectual representations would entirely "depend on our inner activity," as Kant says to Herz, we have no good reason to believe that they conform to an independent intelligible world. Such a priori intellectual representations could well be figments of the brain that do not correspond to anything independent of the human mind. In any case, it is completely mysterious how there might come to be a correspondence between purely intellectual representations and an independent intelligible world.

Kant's strategy in the Critique is similar to that of the Inaugural Dissertation in that both works attempt to reconcile modern science with traditional morality and religion by relegating them to distinct sensible and intelligible worlds, respectively. But the Critique gives a far more modest and yet revolutionary account of a priori knowledge. As Kant's letter to Herz suggests, the main problem with his view in the Inaugural Dissertation is that it tries to explain the possibility of a priori knowledge about a world that is entirely independent of the human mind. This turned out to be a dead end, and Kant never again maintained that we can have a priori knowledge about an intelligible world precisely because such a world would be entirely independent of us. However, Kant's revolutionary position in the Critique is that we can have a priori knowledge about the general structure of the sensible world because it is not entirely independent of the human mind. The sensible world, or the world of appearances, is constructed by the human mind from a combination of sensory matter that we receive passively and a priori forms that are

supplied by our cognitive faculties. We can have a priori knowledge only about aspects of the sensible world that reflect the a priori forms supplied by our cognitive faculties. In Kant's words, "we can cognize of things a priori only what we ourselves have put into them" (Bxviii). So according to the Critique, a priori knowledge is possible only if and to the extent that the sensible world itself depends on the way the human mind structures its experience.

Kant characterizes this new constructivist view of experience in the Critique through an analogy with the revolution wrought by Copernicus in astronomy:

> Up to now it has been assumed that all our cognition must conform to the objects; but all attempts to find out something about them a priori through concepts that would extend our cognition have, on this presupposition, come to nothing. Hence let us once try whether we do not get farther with the problems of metaphysics by assuming that the objects must conform to our cognition, which would agree better with the requested possibility of an a priori cognition of them, which is to establish something about objects before they are given to us. This would be just like the first thoughts of Copernicus, who, when he did not make good progress in the explanation of the celestial motions if he assumed that the entire celestial host revolves around the observer, tried to see if he might not have greater success if he made the observer revolve and left the stars at rest. Now in metaphysics we can try in a similar way regarding the intuition of objects. If intuition has to conform to the constitution of the objects, then I do not see how we can know anything of them a priori; but if the object (as an object of the senses) conforms to the constitution of our faculty of intuition, then I can very well represent this possibility to myself. Yet because I cannot stop with these intuitions, if they are to become

cognitions, but must refer them as representations to something as their object and determine this object through them, I can assume either that the concepts through which I bring about this determination also conform to the objects, and then I am once again in the same difficulty about how I could know anything about them a priori, or else I assume that the objects, or what is the same thing, the experience in which alone they can be cognized (as given objects) conforms to those concepts, in which case I immediately see an easier way out of the difficulty, since experience itself is a kind of cognition requiring the understanding, whose rule I have to presuppose in myself before any object is given to me, hence a priori, which rule is expressed in concepts a priori, to which all objects of experience must therefore necessarily conform, and with which they must agree. (Bxvi–xviii)

As this passage suggests, what Kant has changed in the Critique is primarily his view about the role and powers of the understanding, since he already held in the Inaugural Dissertation that sensibility contributes the forms of space and time – which he calls pure (or a priori) intuitions (2:397) – to our cognition of the sensible world. But the Critique claims that pure understanding too, rather than giving us insight into an intelligible world, is limited to providing forms – which he calls pure or a priori concepts – that structure our cognition of the sensible world. So now both sensibility and understanding work together to construct cognition of the sensible world, which therefore conforms to the a priori forms that are supplied by our cognitive faculties: the a priori intuitions of sensibility and the a priori concepts of the understanding. This account is analogous to the heliocentric revolution of Copernicus in astronomy because both require contributions from the observer to be factored into explanations of phenomena, although neither reduces phenomena to the contributions of observers alone.[6] The way celestial phenomena appear to us on earth, according to Copernicus, is affected by both the motions of celestial

bodies and the motion of the earth, which is not a stationary body around which everything else revolves. For Kant, analogously, the phenomena of human experience depend on both the sensory data that we receive passively through sensibility and the way our mind actively processes this data according to its own a priori rules. These rules supply the general framework in which the sensible world and all the objects (or phenomena) in it appear to us. So the sensible world and its phenomena are not entirely independent of the human mind, which contributes its basic structure.

How does Kant's Copernican revolution in philosophy improve on the strategy of the Inaugural Dissertation for reconciling modern science with traditional morality and religion? First, it gives Kant a new and ingenious way of placing modern science on an a priori foundation. He is now in a position to argue that we can have a priori knowledge about the basic laws of modern science because those laws reflect the human mind's contribution to structuring our experience. In other words, the sensible world necessarily conforms to certain fundamental laws – such as that every event has a cause – because the human mind constructs it according to those laws. Moreover, we can identify those laws by reflecting on the conditions of possible experience, which reveals that it would be impossible for us to experience a world in which, for example, any given event fails to have a cause. From this Kant concludes that metaphysics is indeed possible in the sense that we can have a priori knowledge that the entire sensible world – not just our actual experience, but any possible human experience – necessarily conforms to certain laws. Kant calls this immanent metaphysics or the metaphysics of experience, because it deals with the essential principles that are immanent to human experience.

But, second, if "we can cognize of things a priori only what we ourselves have put into them," then we cannot have a priori knowledge about things whose existence and nature are entirely independent of the human mind, which Kant calls things in themselves (Bxviii). In his words: "[F]rom this

deduction of our faculty of cognizing a priori […] there emerges a very strange result […], namely that with this faculty we can never get beyond the boundaries of possible experience, […and] that such cognition reaches appearances only, leaving the thing in itself as something actual for itself but uncognized by us" (Bxix–xx). That is, Kant's constructivist foundation for scientific knowledge restricts science to the realm of appearances and implies that a priori knowledge of things in themselves that transcend possible human experience – or transcendent metaphysics – is impossible. In the Critique Kant thus rejects the insight into an intelligible world that he defended in the Inaugural Dissertation, and he now claims that rejecting knowledge about things in themselves is necessary for reconciling science with traditional morality and religion. This is because he claims that belief in God, freedom, and immortality have a strictly moral basis, and yet adopting these beliefs on moral grounds would be unjustified if we could know that they were false. "Thus," Kant says, "I had to deny knowledge in order to make room for faith" (Bxxx). Restricting knowledge to appearances and relegating God and the soul to an unknowable realm of things in themselves guarantees that it is impossible to disprove claims about God and the freedom or immortality of the soul, which moral arguments may therefore justify us in believing. Moreover, the determinism of modern science no longer threatens the freedom required by traditional morality, because science and therefore determinism apply only to appearances, and there is room for freedom in the realm of things in themselves, where the self or soul is located. We cannot know (theoretically) that we are free, because we cannot know anything about things in themselves. But there are especially strong moral grounds for the belief in human freedom, which acts as "the keystone" supporting other morally grounded beliefs (5:3–4). In this way, Kant replaces transcendent metaphysics with a new practical science that he calls the metaphysics of morals. It thus turns out that two kinds of metaphysics are possible: the

metaphysics of experience (or nature) and the metaphysics of morals, both of which depend on Kant's Copernican revolution in philosophy.

3. Transcendental idealism

Perhaps the central and most controversial thesis of the Critique of Pure Reason is that human beings experience only appearances, not things in themselves; and that space and time are only subjective forms of human intuition that would not subsist in themselves if one were to abstract from all subjective conditions of human intuition. Kant calls this thesis transcendental idealism.[7] One of his best summaries of it is arguably the following:

> We have therefore wanted to say that all our intuition is nothing but the representation of appearance; that the things that we intuit are not in themselves what we intuit them to be, nor are their relations so constituted in themselves as they appear to us; and that if we remove our own subject or even only the subjective constitution of the senses in general, then all constitution, all relations of objects in space and time, indeed space and time themselves would disappear, and as appearances they cannot exist in themselves, but only in us. What may be the case with objects in themselves and abstracted from all this receptivity of our sensibility remains entirely unknown to us. We are acquainted with nothing except our way of perceiving them, which is peculiar to us, and which therefore does not necessarily pertain to every being, though to be sure it pertains to every human being. We are concerned solely with this. Space and time are its pure forms, sensation in general its matter. We can cognize only the former a priori, i.e., prior to all actual perception, and they are therefore called pure intuition; the latter, however, is that in our cognition that is responsible for its being called a posteriori cognition, i.e.,

empirical intuition. The former adheres to our sensibility absolutely necessarily, whatever sort of sensations we may have; the latter can be very different. (A42/B59–60)[8]

Kant introduces transcendental idealism in the part of the Critique called the Transcendental Aesthetic, and scholars generally agree that for Kant transcendental idealism encompasses at least the following claims:

- In some sense, human beings experience only appearances, not things in themselves.
- Space and time are not things in themselves, or determinations of things in themselves that would remain if one abstracted from all subjective conditions of human intuition. [Kant labels this conclusion a) at A26/B42 and again at A32–33/B49. It is at least a crucial part of what he means by calling space and time transcendentally ideal (A28/B44, A35–36/B52)].
- Space and time are nothing other than the subjective forms of human sensible intuition. [Kant labels this conclusion b) at A26/B42 and again at A33/B49–50].
- Space and time are empirically real, which means that "everything that can come before us externally as an object" is in both space and time, and that our internal intuitions of ourselves are in time (A28/B44, A34–35/B51–51).

But scholars disagree widely on how to interpret these claims, and there is no such thing as the standard interpretation of Kant's transcendental idealism. Two general types of interpretation have been especially influential, however. This section provides an overview of these two interpretations, although it should be emphasized that much important scholarship on transcendental idealism does not fall neatly into either of these two camps.

3.1 The two-objects interpretation

The two-objects reading is the traditional interpretation of Kant's transcendental idealism. It goes back to the earliest review of the Critique – the so-called Göttingen review by Christian Garve (1742–1798) and J. G. Feder (1740–1821)[9] – and it was the dominant way of interpreting Kant's transcendental idealism during his own lifetime. It has been a live interpretive option since then and remains so today, although it no longer enjoys the dominance that it once did.[10]

According to the two-objects interpretation, transcendental idealism is essentially a metaphysical thesis that distinguishes between two classes of objects: appearances and things in themselves. Another name for this view is the two-worlds interpretation, since it can also be expressed by saying that transcendental idealism essentially distinguishes between a world of appearances and another world of things in themselves.

Things in themselves, on this interpretation, are absolutely real in the sense that they would exist and have whatever properties they have even if no human beings were around to perceive them. Appearances, on the other hand, are not absolutely real in that sense, because their existence and properties depend on human perceivers. Moreover, whenever appearances do exist, in some sense they exist in the mind of human perceivers. So appearances are mental entities or mental representations. This, coupled with the claim that we experience only appearances, makes transcendental idealism a form of phenomenalism on this interpretation, because it reduces the objects of experience to mental representations. All of our experiences – all of our perceptions of objects and events in space, even those objects and events themselves, and all non-spatial but still temporal thoughts and feelings – fall into the class of appearances that exist in the mind of human perceivers. These appearances cut us off entirely from the reality of things in themselves, which are non-spatial and non-temporal.

Yet Kant's theory, on this interpretation, nevertheless requires that things in themselves exist, because they must transmit to us the sensory data from which we construct appearances. In principle we cannot know how things in themselves affect our senses, because our experience and knowledge is limited to the world of appearances constructed by and in the mind. Things in themselves are therefore a sort of theoretical posit, whose existence and role are required by the theory but are not directly verifiable.

The main problems with the two-objects interpretation are philosophical. Most readers of Kant who have interpreted his transcendental idealism in this way have been – often very – critical of it, for reasons such as the following:

First, at best Kant is walking a fine line in claiming on the one hand that we can have no knowledge about things in themselves, but on the other hand that we know that things in themselves exist, that they affect our senses, and that they are non-spatial and non-temporal. At worst his theory depends on contradictory claims about what we can and cannot know about things in themselves. This objection was influentially articulated by Jacobi, when he complained that "without that presupposition [of things in themselves] I could not enter into the system, but with it I could not stay within it" (Jacobi 1787, 336).

Second, even if that problem is surmounted, it has seemed to many that Kant's theory, interpreted in this way, implies a radical form of skepticism that traps each of us within the contents of our own mind and cuts us off from reality. Some versions of this objection proceed from premises that Kant rejects. One version maintains that things in themselves are real while appearances are not, and hence that on Kant's view we cannot have experience or knowledge of reality. But Kant denies that appearances are unreal: they are just as real as things in themselves but are in a different metaphysical class. Another version claims that truth always involves a

correspondence between mental representations and things in themselves, from which it would follow that on Kant's view it is impossible for us to have true beliefs about the world. But just as Kant denies that things in themselves are the only (or privileged) reality, he also denies that correspondence with things in themselves is the only kind of truth. Empirical judgments are true just in case they correspond with their empirical objects in accordance with the a priori principles that structure all possible human experience. But the fact that Kant can appeal in this way to an objective criterion of empirical truth that is internal to our experience has not been enough to convince some critics that Kant is innocent of an unacceptable form of skepticism, mainly because of his insistence on our irreparable ignorance about things in themselves.

Third and finally, Kant's denial that things in themselves are spatial or temporal has struck many of his readers as incoherent. The role of things in themselves, on the two-object interpretation, is to affect our senses and thereby to provide the sensory data from which our cognitive faculties construct appearances within the framework of our a priori intuitions of space and time and a priori concepts such as causality. But if there is no space, time, change, or causation in the realm of things in themselves, then how can things in themselves affect us? Transcendental affection seems to involve a causal relation between things in themselves and our sensibility. If this is simply the way we unavoidably think about transcendental affection, because we can give positive content to this thought only by employing the concept of a cause, while it is nevertheless strictly false that things in themselves affect us causally, then it seems not only that we are ignorant of how things in themselves really affect us. It seems, rather, to be incoherent that things in themselves could affect us at all if they are not in space or time.

3.2 The two-aspects interpretation

The two-aspects reading attempts to interpret Kant's transcendental idealism in a way that enables it to be defended against at least some of these objections. On this view, transcendental idealism does not distinguish between two classes of objects but rather between two different aspects of one and the same class of objects. For this reason it is also called the one-world interpretation, since it holds that there is only one world in Kant's ontology, and that at least some objects in that world have two different aspects: one aspect that appears to us, and another aspect that does not appear to us. That is, appearances are aspects of the same objects that also exist in themselves. So, on this reading, appearances are not mental representations, and transcendental idealism is not a form of phenomenalism.[11]

There are at least two main versions of the two-aspects theory. One version treats transcendental idealism as a metaphysical theory according to which objects have two aspects in the sense that they have two sets of properties: one set of relational properties that appear to us and are spatial and temporal, and another set of intrinsic properties that do not appear to us and are not spatial or temporal (Langton 1998). This property-dualist interpretation faces epistemological objections similar to those faced by the two-objects interpretation, because we are in no better position to acquire knowledge about properties that do not appear to us than we are to acquire knowledge about objects that do not appear to us. Moreover, this interpretation also seems to imply that things in themselves are spatial and temporal, since appearances have spatial and temporal properties, and on this view appearances are the same objects as things in themselves. But Kant explicitly denies that space and time are properties of things in themselves.

A second version of the two-aspects theory departs more radically from the traditional two-objects interpretation by denying that transcendental idealism is at bottom a metaphysical theory. Instead, it interprets

transcendental idealism as a fundamentally epistemological theory that distinguishes between two standpoints on the objects of experience: the human standpoint, from which objects are viewed relative to epistemic conditions that are peculiar to human cognitive faculties (namely, the a priori forms of our sensible intuition); and the standpoint of an intuitive intellect, from which the same objects could be known in themselves and independently of any epistemic conditions (Allison 2004). Human beings cannot really take up the latter standpoint but can form only an empty concept of things as they exist in themselves by abstracting from all the content of our experience and leaving only the purely formal thought of an object in general. So transcendental idealism, on this interpretation, is essentially the thesis that we are limited to the human standpoint, and the concept of a thing in itself plays the role of enabling us to chart the boundaries of the human standpoint by stepping beyond them in abstract (but empty) thought.

One criticism of this epistemological version of the two-aspects theory is that it avoids the objections to other interpretations by attributing to Kant a more limited project than the text of the Critique warrants. There are passages that support this reading.[12] But there are also many passages in both editions of the Critique in which Kant describes appearances as representations in the mind and in which his distinction between appearances and things in themselves is given not only epistemological but metaphysical significance.[13] It is unclear whether all of these texts admit of a single, consistent interpretation.

4. The transcendental deduction

The transcendental deduction is the central argument of the Critique of Pure Reason and one of the most complex and difficult texts in the history of philosophy. Given its complexity, there are naturally many different

ways of interpreting the deduction.[14] This brief overview provides one perspective on some of its main ideas.

The transcendental deduction occurs in the part of the Critique called the Analytic of Concepts, which deals with the a priori concepts that, on Kant's view, our understanding uses to construct experience together with the a priori forms of our sensible intuition (space and time), which he discussed in the Transcendental Aesthetic. Kant calls these a priori concepts "categories," and he argues elsewhere (in the so-called metaphysical deduction) that they include such concepts as substance and cause. The goal of the transcendental deduction is to show that we have a priori concepts or categories that are objectively valid, or that apply necessarily to all objects in the world that we experience. To show this, Kant argues that the categories are necessary conditions of experience, or that we could not have experience without the categories. In Kant's words:

> [T]he objective validity of the categories, as a priori concepts, rests on the fact that through them alone is experience possible (as far as the form of thinking is concerned). For they then are related necessarily and a priori to objects of experience, since only by means of them can any object of experience be thought at all.

> The transcendental deduction of all a priori concepts therefore has a principle toward which the entire investigation must be directed, namely this: that they must be recognized as a priori conditions of the possibility of experiences (whether of the intuition that is encountered in them, or of the thinking). Concepts that supply the objective ground of the possibility of experience are necessary just for that reason. (A93–94/B126)

The strategy Kant employs to argue that the categories are conditions of experience is the main source of both the obscurity and the ingenuity of

the transcendental deduction. His strategy is to argue that the categories are necessary specifically for self-consciousness, for which Kant often uses the Leibnizian term "apperception."

4.1 Self-consciousness

One way to approach Kant's argument is to contrast his view of self-consciousness with two alternative views that he rejects. Each of these views, both Kant's and those he rejects, can be seen as offering competing answers the question: what is the source of our sense of an ongoing and invariable self that persists throughout all the changes in our experience?

The first answer to this question that Kant rejects is that self-consciousness arises from some particular content being present in each of one's representations. This material conception of self-consciousness, as we may call it, is loosely suggested by Locke's account of personal identity. According to Locke, "it being the same consciousness that makes a Man be himself to himself, personal Identity depends on that only, whether it be annexed only to one individual Substance, or can be continued in a succession of several Substances" (Essay 2.27.10). What Locke calls "the same consciousness" may be understood as some representational content that is always present in my experience and that both identifies any experience as mine and gives me a sense of a continuous self by virtue of its continual presence in my experience. One problem with this view, Kant believes, is that there is no such representational content that is invariably present in experience, so the sense of an ongoing self cannot possibly arise from that non-existent content (what Locke calls "consciousness") being present in each of one's representations. In Kant's words, self-consciousness "does not yet come about by my accompanying each representation with consciousness, but rather by my adding one representation to the other and being conscious of their synthesis. Therefore it is only because I can combine a manifold of given

representations in one consciousness that it is possible for me to represent the identity of the consciousness in these representations" (B133). Here Kant claims, against the Lockean view, that self-consciousness arises from combining (or synthesizing) representations with one another regardless of their content. In short, Kant has a formal conception of self-consciousness rather than a material one. Since no particular content of my experience is invariable, self-consciousness must derive from my experience having an invariable form or structure, and consciousness of the identity of myself through all of my changing experiences must consist in awareness of the formal unity and law-governed regularity of my experience. The continuous form of my experience is the necessary correlate for my sense of a continuous self.

There are at least two possible versions of the formal conception of self-consciousness: a realist and an idealist version. On the realist version, nature itself is law-governed and we become self-conscious by attending to its law-governed regularities, which also makes this an empiricist view of self-consciousness. The idea of an identical self that persists throughout all of our experience, on this view, arises from the law-governed regularity of nature, and our representations exhibit order and regularity because reality itself is ordered and regular. But Kant rejects this view and embraces a conception of self-consciousness that is both formal and idealist. According to Kant, the formal structure of our experience, its unity and law-governed regularity, is an achievement of our cognitive faculties rather than a property of reality in itself. Our experience has a constant form because our mind constructs experience in a law-governed way. So self-consciousness, for Kant, consists in awareness of the mind's law-governed activity of synthesizing or combining sensible data to construct a unified experience. As he expresses it, "this unity of consciousness would be impossible if in the cognition of the manifold the mind could not become conscious of the identity of the function by means

of which this manifold is synthetically combined into one cognition" (A108).

Kant argues for this formal idealist conception of self-consciousness, and against the formal realist view, on the grounds that "we can represent nothing as combined in the object without having previously combined it ourselves" (B130). In other words, even if reality in itself were law-governed, its laws could not simply migrate over to our mind or imprint themselves on us while our mind is entirely passive. We must exercise an active capacity to represent the world as combined or ordered in a law-governed way, because otherwise we could not represent the world as law-governed even if it were law-governed in itself. Moreover, this capacity to represent the world as law-governed must be a priori because it is a condition of self-consciousness, and we would already have to be self-conscious in order to learn from our experience that there are law-governed regularities in the world. So it is necessary for self-consciousness that we exercise an a priori capacity to represent the world as law-governed. But this would also be sufficient for self-consciousness if we could exercise our a priori capacity to represent the world as law-governed even if reality in itself were not law-governed. In that case, the realist and empiricist conception of self-consciousness would be false, and the formal idealist view would be true.

Kant's confidence that no empiricist account could possibly explain self-consciousness may be based on his assumption that the sense of self each of us has, the thought of oneself as identical throughout all of one's changing experiences, involves necessity and universality, which on his view are the hallmarks of the a priori. This assumption is reflected in what we may call Kant's principle of apperception: "The I think must be able to accompany all my representations; for otherwise something would be represented in me that could not be thought at all, which is as much as to say that the representation would either be impossible or else at least

would be nothing for me" (B131–132).[15] Notice the claims about necessity and universality embodied in the words "must" and "all" here. Kant is saying that for a representation to count as mine, it must necessarily be accessible to conscious awareness in some (perhaps indirect) way: I must be able to accompany it with "I think…." All of my representations must be accessible to consciousness in this way (but they need not actually be conscious), because again that is simply what makes a representation count as mine. Self-consciousness for Kant therefore involves a priori knowledge about the necessary and universal truth expressed in this principle of apperception, and a priori knowledge cannot be based on experience.

Recently it has been argued that Kant developed this thread of his argument in the transcendental deduction after reading Johann Nicolaus Tetens (1736–1807) rather than simply through a direct encounter with Locke's texts (Tetens 1777, Kitcher 2011). On the subject of self-consciousness, Tetens was a follower of Locke and also engaged with Hume's arguments for rejecting a continuing self. So Kant's actual opponents in the deduction may have been Lockean and Humean positions as represented by Tetens, as well as rationalist views that Kant would have encountered directly in texts by Leibniz, Wolff, and some of their followers.

4.2 Objectivity and judgment

On the basis of this formal idealist conception of self-consciousness, Kant's argument (at least one central thread of it) moves through two more conditions of self-consciousness in order to establish the objective validity of the categories. The next condition is that self-consciousness requires me to represent an objective world distinct from my subjective representations – that is, distinct from my thoughts about and sensations of that objective

world. Kant uses this connection between self-consciousness and objectivity to insert the categories into his argument.

In order to be self-conscious, I cannot be wholly absorbed in the contents of my perceptions but must distinguish myself from the rest of the world. But if self-consciousness is an achievement of the mind, then how does the mind achieve this sense that there is a distinction between the I that perceives and the contents of its perceptions? According to Kant, the mind achieves this by distinguishing representations that necessarily belong together from representations that are not necessarily connected but are merely associated in a contingent way. Consider Kant's example of the perception of a house (B162). Imagine a house that is too large to fit into your visual field from your vantage point near its front door. Now imagine that you walk around the house, successively perceiving each of its sides. Eventually you perceive the entire house, but not all at once, and you judge that each of your representations of the sides of the house necessarily belong together (as sides of one house) and that anyone who denied this would be mistaken. But now imagine that you grew up in this house and associate a feeling of nostalgia with it. You would not judge that representations of this house are necessarily connected with feelings of nostalgia. That is, you would not think that other people seeing the house for the first time would be mistaken if they denied that it is connected with nostalgia, because you recognize that this house is connected with nostalgia for you but not necessarily for everyone. Yet you distinguish this merely subjective connection from the objective connection between sides of the house, which is objective because the sides of the house necessarily belong together "in the object," because this connection holds for everyone universally, and because it is possible to be mistaken about it. The point here is not that we must successfully identify which representations necessarily belong together and which are merely associated contingently, but rather that to be self-conscious we must at

least make this general distinction between objective and merely subjective connections of representations.

At this point (at least in the second edition text) Kant introduces the key claim that judgment is what enables us to distinguish objective connections of representations that necessarily belong together from merely subjective and contingent associations: "[A] judgment is nothing other than the way to bring given cognitions to the objective unity of apperception. That is the aim of the copula is in them: to distinguish the objective unity of given representations from the subjective. For this word designates the relation of the representations to the original apperception and its necessary unity" (B141–142). Kant is speaking here about the mental act of judging that results in the formation of a judgment. Judging is an act of what Kant calls synthesis, which he defines as "the action of putting different representations together with each other and comprehending their manifoldness in one cognition" (A77/B103). In other words, to synthesize is in general to combine several representations into a single (more) complex representation, and to judge is specifically to combine concepts into a judgment – that is, to join a subject concept to a predicate concept by means of the copula, as in "the body is heavy" or "the house is four-sided." Judgments need not be true, of course, but they always have a truth value (true or false) because they make claims to objective validity. When I say, by contrast, that "If I carry a body, I feel a pressure of weight," or that "if I see this house, I feel nostalgia," I am not making a judgment about the object (the body or the house) but rather I am expressing a subjective association that may apply only to me (B142). [16]

Kant's reference to the necessary unity of apperception or self-consciousness in the quotation above means (at least) that the action of judging is the way our mind achieves self-consciousness. We must represent an objective world in order to distinguish ourselves from it, and

we represent an objective world by judging that some representations necessarily belong together. Moreover, recall from 4.1 that, for Kant, we must have an a priori capacity to represent the world as law-governed, because "we can represent nothing as combined (or connected) in the object without having previously combined it ourselves" (B130). It follows that objective connections in the world cannot simply imprint themselves on our mind. Rather, experience of an objective world must be constructed by exercising an a priori capacity to judge, which Kant calls the faculty of understanding (A80–81/B106). The understanding constructs experience by providing the a priori rules, or the framework of necessary laws, in accordance with which we judge representations to be objective. These rules are the pure concepts of the understanding or categories, which are therefore conditions of self-consciousness, since they are rules for judging about an objective world, and self-consciousness requires that we distinguish ourselves from an objective world.

Kant identifies the categories in what he calls the metaphysical deduction, which precedes the transcendental deduction.[17] Very briefly, since the categories are a priori rules for judging, Kant argues that an exhaustive table of categories can be derived from a table of the basic logical forms of judgments. For example, according to Kant the logical form of the judgment that "the body is heavy" would be singular, affirmative, categorical, and assertoric. But since categories are not mere logical functions but instead are rules for making judgments about objects or an objective world, Kant arrives at his table of categories by considering how each logical function would structure judgments about objects (within our spatio-temporal forms of intuition). For example, he claims that categorical judgments express a logical relation between subject and predicate that corresponds to the ontological relation between substance and accident; and the logical form of a hypothetical judgment expresses a relation that corresponds to cause and effect. Taken together with this argument, then, the transcendental deduction argues that we become self-

conscious by representing an objective world of substances that interact according to causal laws.

4.3 The law-giver of nature

The final condition of self-consciousness that Kant adds to the preceding conditions is that our understanding must cooperate with sensibility to construct one, unbounded, and unified space-time to which all of our representations may be related.

To see why this further condition is required, consider that so far we have seen why Kant holds that we must represent an objective world in order to be self-conscious, but we could represent an objective world even if it were not possible to relate all of our representations to this objective world. For all that has been said so far, we might still have unruly representations that we cannot relate in any way to the objective framework of our experience. On Kant's view, this would be a problem because, as we have seen, he holds that self-consciousness involves universality and necessity: according to his principle of apperception, "the I think must be able to accompany all my representations" (B131). Yet if, on the one hand, I had representations that I could not relate in some way to an objective world, then I could not accompany those representations with "I think" or recognize them as my representations, because I can say "I think…" about any given representation only by relating it to an objective world, according to the argument just discussed. So I must be able to relate any given representation to an objective world in order for it to count as mine. On the other hand, self-consciousness would also be impossible if I represented multiple objective worlds, even if I could relate all of my representations to some objective world or other. In that case, I could not become conscious of an identical self that has, say, representation 1 in space-time A and representation 2 in space-time B. It may be possible to imagine disjointed spaces and times, but it is not

possible to represent them as objectively real. So self-consciousness requires that I can relate all of my representations to a single objective world.

The reason why I must represent this one objective world by means of a unified and unbounded space-time is that, as Kant argued in the Transcendental Aesthetic, space and time are the pure forms of human intuition. If we had different forms of intuition, then our experience would still have to constitute a unified whole in order for us to be self-conscious, but this would not be a spatio-temporal whole. Given that space and time are our forms of intuition, however, our understanding must still cooperate with sensibility to construct a spatio-temporal whole of experience because, once again, "we can represent nothing as combined in the object without having previously combined it ourselves," and "all combination [...] is an action of the understanding" (B130). So Kant distinguishes between space and time as pure forms of intuition, which belong solely to sensibility; and the formal intuitions of space and time (or space-time), which are unified by the understanding (B160–161). These formal intuitions are the spatio-temporal whole within which our understanding constructs experience in accordance with the categories.[18]

The most important implication of Kant's claim that the understanding constructs a single whole of experience to which all of our representations can be related is that, since he defines nature "regarded materially" as "the sum total of all appearances" and he has argued that the categories are objectively valid of all possible appearances, on his view it follows that our categories are the source of the fundamental laws of nature "regarded formally" (B163, 165). So Kant concludes on this basis that the understanding is the true law-giver of nature. In his words: "all appearances in nature, as far as their combination is concerned, stand under the categories, on which nature (considered merely as nature in general) depends, as the original ground of its necessary lawfulness (as

nature regarded formally)" (ibid.). Or more strongly: "we ourselves bring into the appearances that order and regularity that we call nature, and moreover we would not be able to find it there if we, or the nature of our mind, had not originally put it there. […] The understanding is thus not merely a faculty for making rules through the comparison of the appearances: it is itself the legislation for nature, i.e., without understanding there would not be any nature at all" (A125–126).

5. Morality and freedom

Having examined two central parts of Kant's positive project in theoretical philosophy from the Critique of Pure Reason, transcendental idealism and the transcendental deduction, let us now turn to his practical philosophy in the Critique of Practical Reason. Since Kant's philosophy is deeply systematic, this section begins with a preliminary look at how his theoretical and practical philosophy fit together (see also section 7).

5.1 Theoretical and practical autonomy

The fundamental idea of Kant's philosophy is human autonomy. So far we have seen this in Kant's constructivist view of experience, according to which our understanding is the source of the general laws of nature. "Autonomy" literally means giving the law to oneself, and on Kant's view our understanding provides laws that constitute the a priori framework of our experience. Our understanding does not provide the matter or content of our experience, but it does provide the basic formal structure within which we experience any matter received through our senses. Kant's central argument for this view is the transcendental deduction, according to which it is a condition of self-consciousness that our understanding constructs experience in this way. So we may call self-consciousness the highest principle of Kant's theoretical philosophy, since it is (at least) the basis for all of our a priori knowledge about the structure of nature.

Kant's moral philosophy is also based on the idea of autonomy. He holds that there is a single fundamental principle of morality, on which all specific moral duties are based. He calls this moral law (as it is manifested to us) the categorical imperative (see 5.4). The moral law is a product of reason, for Kant, while the basic laws of nature are products of our understanding. There are important differences between the senses in which we are autonomous in constructing our experience and in morality. For example, Kant regards understanding and reason as different cognitive faculties, although he sometimes uses "reason" in a wide sense to cover both.[19] The categories and therefore the laws of nature are dependent on our specifically human forms of intuition, while reason is not. The moral law does not depend on any qualities that are peculiar to human nature but only on the nature of reason as such, although its manifestation to us as a categorical imperative (as a law of duty) reflects the fact that the human will is not necessarily determined by pure reason but is also influenced by other incentives rooted in our needs and inclinations; and our specific duties deriving from the categorical imperative do reflect human nature and the contingencies of human life. Despite these differences, however, Kant holds that we give the moral law to ourselves, just as we also give the general laws of nature to ourselves, though in a different sense. Moreover, we each necessarily give the same moral law to ourselves, just as we each construct our experience in accordance with the same categories. To summarize:

- Theoretical philosophy is about how the world is (A633/B661). Its highest principle is self-consciousness, on which our knowledge of the basic laws of nature is based. Given sensory data, our understanding constructs experience according to these a priori laws.
- Practical philosophy is about how the world ought to be (ibid., A800–801/B828–829). Its highest principle is the moral law, from which we derive duties that command how we ought to act in specific situations. Kant also claims that reflection on our moral duties and

our need for happiness leads to the thought of an ideal world, which he calls the highest good (see section 6). Given how the world is (theoretical philosophy) and how it ought to be (practical philosophy), we aim to make the world better by constructing or realizing the highest good.

So both parts of Kant's philosophy are about autonomously constructing a world, but in different senses. In theoretical philosophy, we use our categories and forms of intuition to construct a world of experience or nature. In practical philosophy, we use the moral law to construct the idea of a moral world or a realm of ends that guides our conduct (4:433), and ultimately to transform the natural world into the highest good. Finally, transcendental idealism is the framework within which these two parts of Kant's philosophy fit together (20:311). Theoretical philosophy deals with appearances, to which our knowledge is strictly limited; and practical philosophy deals with things in themselves, although it does not give us knowledge about things in themselves but only provides rational justification for certain beliefs about them for practical purposes.

To understand Kant's arguments that practical philosophy justifies certain beliefs about things in themselves, it is necessary to see them in the context of his criticism of German rationalist metaphysics. The three traditional topics of Leibniz-Wolffian special metaphysics were rational psychology, rational cosmology, and rational theology, which dealt, respectively, with the human soul, the world-whole, and God. In the part of the Critique of Pure Reason called the Transcendental Dialectic, Kant argues against the Leibniz-Wolffian view that human beings are capable of a priori knowledge in each of these domains, and he claims that the errors of Leibniz-Wolffian metaphysics are due to an illusion that has its seat in the nature of human reason itself. According to Kant, human reason necessarily produces ideas of the soul, the world-whole, and God; and these ideas unavoidably produce the illusion that we have a priori

knowledge about transcendent objects corresponding to them. This is an illusion, however, because in fact we are not capable of a priori knowledge about any such transcendent objects. Nevertheless, Kant attempts to show that these illusory ideas have a positive, practical use. He thus reframes Leibniz-Wolffian special metaphysics as a practical science that he calls the metaphysics of morals. On Kant's view, our ideas of the soul, the world-whole, and God provide the content of morally justified beliefs about human immortality, human freedom, and the existence of God, respectively; but they are not proper objects of speculative knowledge.[20]

5.2 Freedom

The most important belief about things in themselves that Kant thinks only practical philosophy can justify concerns human freedom. Freedom is important because, on Kant's view, moral appraisal presupposes that we are free in the sense that we have the ability to do otherwise. To see why, consider Kant's example of a man who commits a theft (5:95ff.). Kant holds that in order for this man's action to be morally wrong, it must have been within his control in the sense that it was within his power at the time not to have committed the theft. If this was not within his control at the time, then, while it may be useful to punish him in order to shape his behavior or to influence others, it nevertheless would not be correct to say that his action was morally wrong. Moral rightness and wrongness apply only to free agents who control their actions and have it in their power, at the time of their actions, either to act rightly or not. According to Kant, this is just common sense.

On these grounds, Kant rejects a type of compatibilism that he calls the "comparative concept of freedom" and associates with Leibniz (5:96–97). (Note that Kant has a specific type of compatibilism in mind, which I will refer to simply as "compatibilism," although there may be other types of compatibilism that do not fit Kant's characterization of that view). On the

compatibilist view, as Kant understands it, I am free whenever the cause of my action is within me. So I am unfree only when something external to me pushes or moves me, but I am free whenever the proximate cause of my body's movement is internal to me as an "acting being" (5:96). If we distinguish between involuntary convulsions and voluntary bodily movements, then on this view free actions are just voluntary bodily movements. Kant ridicules this view as a "wretched subterfuge" that tries to solve an ancient philosophical problem "with a little quibbling about words" (ibid.). This view, he says, assimilates human freedom to "the freedom of a turnspit," or a projectile in flight, or the motion of a clock's hands (5:96–97). The proximate causes of these movements are internal to the turnspit, the projectile, and the clock at the time of the movement. This cannot be sufficient for moral responsibility.

Why not? The reason, Kant says, is ultimately that the causes of these movements occur in time. Return to the theft example. A compatibilist would say that the thief's action is free because its proximate cause is inside him, and because the theft was not an involuntary convulsion but a voluntary action. The thief decided to commit the theft, and his action flowed from this decision. According to Kant, however, if the thief's decision is a natural phenomenon that occurs in time, then it must be the effect of some cause that occurred in a previous time. This is an essential part of Kant's Newtonian worldview and is grounded in the a priori laws (specifically, the category of cause and effect) in accordance with which our understanding constructs experience: every event has a cause that begins in an earlier time. If that cause too was an event occurring in time, then it must also have a cause beginning in a still earlier time, etc. All natural events occur in time and are thoroughly determined by causal chains that stretch backwards into the distant past. So there is no room for freedom in nature, which is deterministic in a strong sense.

The root of the problem, for Kant, is time. Again, if the thief's choice to commit the theft is a natural event in time, then it is the effect of a causal chain extending into the distant past. But the past is out of his control now, in the present. Once the past is past, he can't change it. On Kant's view, that is why his actions would not be in his control in the present if they are determined by events in the past. Even if he could control those past events in the past, he cannot control them now. But in fact past events were not in his control in the past either if they too were determined by events in the more distant past, because eventually the causal antecedents of his action stretch back before his birth, and obviously events that occurred before his birth were not in his control. So if the thief's choice to commit the theft is a natural event in time, then it is not now and never was in his control, and he could not have done otherwise than to commit the theft. In that case, it would be a mistake to hold him morally responsible for it.

Compatibilism, as Kant understands it, therefore locates the issue in the wrong place. Even if the cause of my action is internal to me, if it is in the past – for example, if my action today is determined by a decision I made yesterday, or from the character I developed in childhood – then it is not within my control now. The real issue is not whether the cause of my action is internal or external to me, but whether it is in my control now. For Kant, however, the cause of my action can be within my control now only if it is not in time. This is why Kant thinks that transcendental idealism is the only way to make sense of the kind of freedom that morality requires. For transcendental idealism allows that the cause of my action may be a thing in itself outside of time: namely, my noumenal self, which is free because it is not part of nature. No matter what kind of character I have developed or what external influences act on me, on Kant's view all of my intentional, voluntary actions are immediate effects of my noumenal self, which is causally undetermined (5:97–98). My noumenal self is an uncaused cause outside of time, which therefore is not

subject to the deterministic laws of nature in accordance with which our understanding constructs experience.

Many puzzles arise on this picture that Kant does not resolve. For example, if my understanding constructs all appearances in my experience of nature, not only appearances of my own actions, then why am I responsible only for my own actions but not for everything that happens in the natural world? Moreover, if I am not alone in the world but there are many noumenal selves acting freely and incorporating their free actions into the experience they construct, then how do multiple transcendentally free agents interact? How do you integrate my free actions into the experience that your understanding constructs?[21] In spite of these unsolved puzzles, Kant holds that we can make sense of moral appraisal and responsibility only by thinking about human freedom in this way, because it is the only way to prevent natural necessity from undermining both.

Finally, since Kant invokes transcendental idealism to make sense of freedom, interpreting his thinking about freedom leads us back to disputes between the two-objects and two-aspects interpretations of transcendental idealism. On the face of it, the two-objects interpretation seems to make better sense of Kant's view of transcendental freedom than the two-aspects interpretation. If morality requires that I am transcendentally free, then it seems that my true self, and not just an aspect of my self, must be outside of time, according to Kant's argument. But applying the two-objects interpretation to freedom raises problems of its own, since it involves making a distinction between noumenal and phenomenal selves that does not arise on the two-aspects view. If only my noumenal self is free, and freedom is required for moral responsibility, then my phenomenal self is not morally responsible. But how are my noumenal and phenomenal selves related, and why is punishment inflicted on phenomenal selves? It is unclear whether and to what extent appealing to Kant's theory of freedom

can help to settle disputes about the proper interpretation of transcendental idealism, since there are serious questions about the coherence of Kant's theory on either interpretation.

5.3 The fact of reason

Can we know that we are free in this transcendental sense? Kant's response is tricky. On the one hand, he distinguishes between theoretical knowledge and morally justified belief (A820–831/B848–859). We do not have theoretical knowledge that we are free or about anything beyond the limits of possible experience, but we are morally justified in believing that we are free in this sense. On the other hand, Kant also uses stronger language than this when discussing freedom. For example, he says that "among all the ideas of speculative reason freedom is the only one the possibility of which we know a priori, though without having any insight into it, because it is the condition of the moral law, which we do know." In a footnote to this passage, Kant explains that we know freedom a priori because "were there no freedom, the moral law would not be encountered at all in ourselves," and on Kant's view everyone does encounter the moral law a priori (5:4). For this reason, Kant claims that the moral law "proves" the objective, "though only practical, undoubted reality" of freedom (5:48–49). So Kant wants to say that we do have knowledge of the reality of freedom, but that this is practical knowledge of a practical reality, or cognition "only for practical purposes," by which he means to distinguish it from theoretical knowledge based on experience or reflection on the conditions of experience (5:133). Our practical knowledge of freedom is based instead on the moral law. The difference between Kant's stronger and weaker language seems mainly to be that his stronger language emphasizes that our belief or practical knowledge about freedom is unshakeable and that it in turn provides support for other morally grounded beliefs in God and the immortality of the soul.

Kant calls our consciousness of the moral law, our awareness that the moral law binds us or has authority over us, the "fact of reason" (5:31–32, 42–43, 47, 55). So, on his view, the fact of reason is the practical basis for our belief or practical knowledge that we are free. Kant insists that this moral consciousness is "undeniable," "a priori," and "unavoidable" (5:32, 47, 55). Every human being has a conscience, a common sense grasp of morality, and a firm conviction that he or she is morally accountable. We may have different beliefs about the source of morality's authority – God, social convention, human reason. We may arrive at different conclusions about what morality requires in specific situations. And we may violate our own sense of duty. But we all have a conscience, and an unshakeable belief that morality applies to us. According to Kant, this belief cannot and does not need to be justified or "proved by any deduction" (5:47). It is just a ground-level fact about human beings that we hold ourselves morally accountable. But Kant is making a normative claim here as well: it is also a fact, which cannot and does not need to be justified, that we are morally accountable, that morality does have authority over us. Kant holds that philosophy should be in the business of defending this common sense moral belief, and that in any case it could never prove or disprove it (4:459).

Kant may hold that the fact of reason, or our consciousness of moral obligation, implies that we are free on the grounds that ought implies can. In other words, Kant may believe that it follows from the fact that we ought (morally) to do something that we can or are able to do it. This is suggested, for example, by a passage in which Kant asks us to imagine someone threatened by his prince with immediate execution unless he "give[s] false testimony against an honorable man whom the prince would like to destroy under a plausible pretext." Kant says that "[h]e would perhaps not venture to assert whether he would do it or not, but he must admit without hesitation that it would be possible for him. He judges, therefore, that he can do something because he is aware that he ought to

do it and cognizes freedom within him, which, without the moral law, would have remained unknown to him" (5:30). This is a hypothetical example of an action not yet carried out. It seems that pangs of guilt about the immorality of an action that you carried out in the past, on this reasoning, would imply more directly that you have (or at least had) the ability to act otherwise than you did, and therefore that you are free in Kant's sense.

5.4 The categorical imperative

In both the Groundwork of the Metaphysics of Morals and the Critique of Practical Reason, Kant also gives a more detailed argument for the conclusion that morality and freedom reciprocally imply one another, which is sometimes called the reciprocity thesis (Allison 1990). On this view, to act morally is to exercise freedom, and the only way to fully exercise freedom is to act morally. Kant's arguments for this view differ in these texts, but the general structure of his argument in the Critique of Practical Reason may be summarized as follows.

First, it follows from the basic idea of having a will that to act at all is to act on some principle, or what Kant calls a maxim. A maxim is a subjective rule or policy of action: it says what you are doing and why. Kant gives as examples the maxims "to let no insult pass unavenged" and "to increase my wealth by every safe means" (5:19, 27). We may be unaware of our maxims, we may not act consistently on the same maxims, and our maxims may not be consistent with one another. But Kant holds that since we are rational beings our actions always aim at some sort of end or goal, which our maxim expresses. The goal of an action may be something as basic as gratifying a desire, or it may be something more complex such as becoming a doctor or a lawyer. In any case, the causes of

our actions are never our desires or impulses, on Kant's view. If I act to gratify some desire, then I choose to act on a maxim that specifies the gratification of that desire as the goal of my action. For example, if I desire some coffee, then I may act on the maxim to go to a cafe and buy some coffee in order to gratify that desire.

Second, Kant distinguishes between two basic kinds of principles or rules that we can act on: what he calls material and formal principles. To act in order to satisfy some desire, as when I act on the maxim to go for coffee at a cafe, is to act on a material principle (5:21ff.). Here the desire (for coffee) fixes the goal, which Kant calls the object or matter of the action, and the principle says how to achieve that goal (go to a cafe). Corresponding to material principles, on Kant's view, are what he calls hypothetical imperatives. A hypothetical imperative is a principle of rationality that says that I should act in a certain way if I choose to satisfy some desire. If maxims in general are rules that describe how one does act, then imperatives in general prescribe how one should act. An imperative is hypothetical if it says how I should act only if I choose to pursue some goal in order to gratify a desire (5:20). This, for example, is a hypothetical imperative: if you want coffee, then go to the cafe. This hypothetical imperative applies to you only if you desire coffee and choose to gratify that desire.

In contrast to material principles, formal principles describe how one acts without making reference to any desires. This is easiest to understand through the corresponding kind of imperative, which Kant calls a categorical imperative. A categorical imperative commands unconditionally that I should act in some way. So while hypothetical imperatives apply to me only on the condition that I have and set the goal of satisfying the desires that they tell me how to satisfy, categorical imperatives apply to me no matter what my goals and desires may be. Kant regards moral laws as categorical imperatives, which apply to

everyone unconditionally. For example, the moral requirement to help others in need does not apply to me only if I desire to help others in need, and the duty not to steal is not suspended if I have some desire that I could satisfy by stealing. Moral laws do not have such conditions but rather apply unconditionally. That is why they apply to everyone in the same way.

Third, insofar as I act only on material principles or hypothetical imperatives, I do not act freely, but rather I act only to satisfy some desire(s) that I have, and what I desire is not ultimately within my control. To some limited extent we are capable of rationally shaping our desires, but insofar as we choose to act in order to satisfy desires we are choosing to let nature govern us rather than governing ourselves (5:118). We are always free in the sense that we always have the capacity to govern ourselves rationally instead of letting our desires set our ends for us. But we may (freely) fail to exercise that capacity. Moreover, since Kant holds that desires never cause us to act, but rather we always choose to act on a maxim even when that maxim specifies the satisfaction of a desire as the goal of our action, it also follows that we are always free in the sense that we freely choose our maxims. Nevertheless, our actions are not free in the sense of being autonomous if we choose to act only on material principles, because in that case we do not give the law to ourselves, but instead we choose to allow nature in us (our desires) to determine the law for our actions.

Finally, the only way to act freely in the full sense of exercising autonomy is therefore to act on formal principles or categorical imperatives, which is also to act morally. Kant does not mean that acting autonomously requires that we take no account of our desires, because that would be impossible (5:25, 61). Rather, he holds that we typically formulate maxims with a view to satisfying our desires, but that "as soon as we draw up maxims of the will for ourselves" we become immediately conscious of the moral law

(5:29). This immediate consciousness of the moral law takes the following form:

> I have, for example, made it my maxim to increase my wealth by every safe means. Now I have a deposit in my hands, the owner of which has died and left no record of it. This is, naturally, a case for my maxim. Now I want only to know whether that maxim could also hold as a universal practical law. I therefore apply the maxim to the present case and ask whether it could indeed take the form of a law, and consequently whether I could through my maxim at the same time give such a law as this: that everyone may deny a deposit which no one can prove has been made. I at once become aware that such a principle, as a law, would annihilate itself since it would bring it about that there would be no deposits at all. (5:27)

In other words, to assess the moral permissibility of my maxim, I ask whether everyone could act on it, or whether it could be willed as a universal law. The issue is not whether it would be good if everyone acted on my maxim, or whether I would like it, but only whether it would be possible for my maxim to be willed as a universal law. This gets at the form, not the matter or content, of the maxim. A maxim has morally permissible form, for Kant, only if it could be willed as a universal law. If my maxim fails this test, as this one does, then it is morally impermissible for me to act on it.

If my maxim passes the universal law test, then it is morally permissible for me to act on it, but I fully exercise my autonomy only if my fundamental reason for acting on this maxim is that it is morally permissible or required that I do so. Imagine that I am moved by a feeling of sympathy to formulate the maxim to help someone in need. In this case, my original reason for formulating this maxim is that a certain feeling moved me. Such feelings are not entirely within my control and may not

be present when someone actually needs my help. But this maxim passes Kant's test: it could be willed as a universal law that everyone help others in need from motives of sympathy. So it would not be wrong to act on this maxim when the feeling of sympathy so moves me. But helping others in need would not fully exercise my autonomy unless my fundamental reason for doing so is not that I have some feeling or desire, but rather that it would be right or at least permissible to do so. Only when such a purely formal principle supplies the fundamental motive for my action do I act autonomously.

So the moral law is a law of autonomy in the sense that "freedom and unconditional practical law reciprocally imply each another" (5:29). Even when my maxims are originally suggested by my feelings and desires, if I act only on morally permissible (or required) maxims because they are morally permissible (or required), then my actions will be autonomous. And the reverse is true as well: for Kant this is the only way to act autonomously.[22]

6. The highest good and practical postulates

Kant holds that reason unavoidably produces not only consciousness of the moral law but also the idea of a world in which there is both complete virtue and complete happiness, which he calls the highest good. Our duty to promote the highest good, on Kant's view, is the sum of all moral duties, and we can fulfill this duty only if we believe that the highest good is a possible state of affairs. Furthermore, we can believe that the highest good is possible only if we also believe in the immortality of the soul and the existence of God, according to Kant. On this basis, he claims that it is morally necessary to believe in the immortality of the soul and the existence of God, which he calls postulates of pure practical reason. This section briefly outlines Kant's view of the highest good and his argument

for these practical postulates in the Critique of Practical Reason and other works.

6.1 The highest good

In the previous section we saw that, on Kant's view, the moral law is a purely formal principle that commands us to act only on maxims that have what he calls lawgiving form, which maxims have only if they can be willed as universal laws. Moreover, our fundamental reason for choosing to act on such maxims should be that they have this lawgiving form, rather than that acting on them would achieve some end or goal that would satisfy a desire (5:27). For example, I should help others in need not, at bottom, because doing so would make me feel good, even if it would, but rather because it is right; and it is right (or permissible) to help others in need because this maxim can be willed as a universal law.

But although Kant holds that the morality of an action depends on the form of its maxim rather than its end or goal, he nevertheless claims both that every human action has an end and that we are unavoidably concerned with the consequences of our actions (4:437; 5:34; 6:5–7, 385). This is not a moral requirement but simply part of what it means to be a rational being. Moreover, Kant also holds the stronger view that it is an unavoidable feature of human reason that we form ideas not only about the immediate and near-term consequences of our actions, but also about ultimate consequences. This is the practical manifestation of reason's general demand for what Kant calls "the unconditioned" (5:107–108).[23] In particular, since we naturally have desires and inclinations, and our reason has "a commission" to attend to the satisfaction of our desires and inclinations, on Kant's view we unavoidably form an idea of the maximal satisfaction of all our inclinations and desires, which he calls happiness (5:61, 22, 124). This idea is indeterminate, however, since nobody can know "what he really wishes and wills" and thus what would make him

completely happy (4:418). We also form the idea of a moral world or realm of ends, in which everyone acts only in accordance with maxims that can be universal laws (A808/B836, 4:433ff.).

But neither of these ideas by itself expresses our unconditionally complete end, as human reason demands in its practical use. A perfectly moral world by itself would not constitute our "whole and complete good […] even in the judgment of an impartial reason," because it is human nature also to need happiness (5:110, 25). And happiness by itself would not be unconditionally good, because moral virtue is a condition of worthiness to be happy (5:111). So our unconditionally complete end must combine both virtue and happiness. In Kant's words, "virtue and happiness together constitute possession of the highest good in a person, and happiness distributed in exact proportion to morality (as the worth of a person and his worthiness to be happy) constitutes the highest good of a possible world" (5:110–111). It is this ideal world combining complete virtue with complete happiness that Kant normally has in mind when he discusses the highest good.

Kant says that we have a duty to promote the highest good, taken in this sense (5:125). He does not mean, however, to be identifying some new duty that is not derived from the moral law, in addition to all the particular duties we have that are derived from the moral law.[24] For example, he is not claiming that in addition to my duties to help others in need, not to commit theft, etc., I also have the additional duty to represent the highest good as the final end of all moral conduct, combined with happiness, and to promote that end. Rather, as we have seen, Kant holds that it is an unavoidable feature of human reasoning, instead of a moral requirement, that we represent all particular duties as leading toward the promotion of the highest good. So the duty to promote the highest good is not a particular duty at all, but the sum of all our duties derived from the moral law – it "does not increase the number of morality's duties but rather

provides these with a special point of reference for the unification of all ends" (6:5). Nor does Kant mean that anyone has a duty to realize or actually bring about the highest good through their own power, although his language sometimes suggests this (5:113, 122). Rather, at least in his later works Kant claims that only the common striving of an entire "ethical community" can actually produce the highest good, and that the duty of individuals is to promote (but not single-handedly produce) this end with all of their strength by doing what the moral law commands (6:97–98, 390–394).[25]

Finally, according to Kant we must conceive of the highest good as a possible state of affairs in order to fulfill our duty to promote it. Here Kant does not mean that we unavoidably represent the highest good as possible, since his view is that we must represent it as possible only if we are to do our duty of promoting it, and yet we may fail at doing our duty. Rather, we have a choice about whether to conceive of the highest good as possible, to regard it as impossible, or to remain noncommittal (5:144–145). But we can fulfill our duty of promoting the highest good only by choosing to conceive of the highest good as possible, because we cannot promote any end without believing that it is possible to achieve that end (5:122). So fulfilling the sum of all moral duties to promote the highest good requires believing that a world of complete virtue and happiness is not simply "a phantom of the mind" but could actually be realized (5:472).

6.2 The postulates of pure practical reason

Kant argues that we can comply with our duty to promote the highest good only if we believe in the immortality of the soul and the existence of God. This is because to comply with that duty we must believe that the highest good is possible, and yet to believe that the highest good is possible we must believe that the soul is immortal and that God exists, according to Kant.[26]

Consider first Kant's moral argument for belief in immortality. The highest good, as we have seen, would be a world of complete morality and happiness. But Kant holds that it is impossible for "a rational being of the sensible world" to exhibit "complete conformity of dispositions with the moral law," which he calls "holiness," because we can never extirpate the propensity of our reason to give priority to the incentives of inclination over the incentive of duty, which propensity Kant calls radical evil (5:122, 6:37). But Kant claims that the moral law nevertheless requires holiness, and that it therefore "can only be found in an endless progress toward that complete conformity," or progress that goes to infinity (5:122). This does not mean that we can substitute endless progress toward complete conformity with the moral law for holiness in the concept of the highest good, but rather that we must represent that complete conformity as an infinite progress toward the limit of holiness. Kant continues: "This endless progress is, however, possible only on the presupposition of the existence and personality of the same rational being continuing endlessly (which is called the immortality of the soul). Hence the highest good is practically possible only on the presupposition of the immortality of the soul, so that this, as inseparable with the moral law, is a postulate of pure practical reason" (ibid.). Kant's idea is not that we should imagine ourselves attaining holiness later although we are not capable of it in this life. Rather, his view is that we must represent holiness as continual progress toward complete conformity of our dispositions with the moral law that begins in this life and extends into infinity.

Kant's moral argument for belief in God in the Critique of Practical Reason may be summarized as follows. Kant holds that virtue and happiness are not just combined but necessarily combined in the idea of the highest good, because only possessing virtue makes one worthy of happiness – a claim that Kant seems to regard as part of the content of the moral law (4:393; 5:110, 124). But we can represent virtue and happiness as necessarily combined only by representing virtue as the efficient cause

of happiness. This means that we must represent the highest good not simply as a state of affairs in which everyone is both happy and virtuous, but rather as one in which everyone is happy because they are virtuous (5:113–114, 124). However, it is beyond the power of human beings, both individually and collectively, to guarantee that happiness results from virtue, and we do not know any law of nature that guarantees this either. Therefore, we must conclude that the highest good is impossible, unless we postulate "the existence of a cause of nature, distinct from nature, which contains the ground of this connection, namely the exact correspondence of happiness with morality" (5:125). This cause of nature would have to be God since it must have both understanding and will. Kant probably does not conceive of God as the efficient cause of a happiness that is rewarded in a future life to those who are virtuous in this one. Rather, his view is probably that we represent our endless progress toward holiness, beginning with this life and extending into infinity, as the efficient cause of our happiness, which likewise begins in this life and extends to a future one, in accordance with teleological laws that God authors and causes to harmonize with efficient causes in nature (A809–812/B837–840; 5:127–131, 447–450).

Both of these arguments are subjective in the sense that, rather than attempting to show how the world must be constituted objectively in order for the highest good to be possible, they purport to show only how we must conceive of the highest good in order to be subjectively capable both of representing it as possible and of fulfilling our duty to promote it. But Kant also claims that both arguments have an objective basis: first, in the sense that it cannot be proven objectively either that immortality or God's existence are impossible; and, second, in the sense that both arguments proceed from a duty to promote the highest good that is based not on the subjective character of human reason but on the moral law, which is objectively valid for all rational beings. So while it is not, strictly speaking, a duty to believe in God or immortality, we must believe both in

order to fulfill our duty to promote the highest good, given the subjective character of human reason.

To see why, consider what would happen if we did not believe in God or immortality, according to Kant. In the Critique of Pure Reason, Kant seems to say that this would leave us without any incentive to be moral, and even that the moral law would be invalid without God and immortality (A813/B841, A468/B496). But Kant later rejects this view (8:139). His mature view is that our reason would be in conflict with itself if we did not believe in God and immortality, because pure practical reason would represent the moral law as authoritative for us and so present us with an incentive that is sufficient to determine our will; but pure theoretical (i.e., speculative) reason would undermine this incentive by declaring morality an empty ideal, since it would not be able to conceive of the highest good as possible (5:121, 143, 471–472, 450–453). In other words, the moral law would remain valid and provide any rational being with sufficient incentive to act from duty, but we would be incapable of acting as rational beings, since "it is a condition of having reason at all […] that its principles and affirmations must not contradict one another" (5:120). The only way to bring speculative and practical reason "into that relation of equality in which reason in general can be used purposively" is to affirm the postulates on the grounds that pure practical reason has primacy over speculative reason. This means, Kant explains, that if the capacity of speculative reason "does not extend to establishing certain propositions affirmatively, although they do not contradict it, as soon as these same propositions belong inseparably to the practical interest of pure reason it must accept them […,] being mindful, however, that these are not its insights but are yet extensions of its use from another, namely a practical perspective" (5:121). The primacy of practical reason is a key element of Kant's response to the crisis of the Enlightenment, since he holds that reason deserves the sovereign authority entrusted to it by the Enlightenment only on this basis.

7. The unity of nature and freedom

This final section briefly discusses how Kant attempts to unify the theoretical and practical parts of his philosophical system in the Critique of the Power of Judgment.

7.1 The great chasm

In the Preface and Introduction to the Critique of the Power of Judgment, Kant announces that his goal in the work is to "bring [his] entire critical enterprise to an end" by bridging the "gulf" or "chasm" that separates the domain of his theoretical philosophy (discussed mainly in the Critique of Pure Reason) from the domain of his practical philosophy (discussed mainly in the Critique of Practical Reason) (5:170, 176, 195). In his words: "The understanding legislates a priori for nature, as object of the senses, for a theoretical cognition of it in a possible experience. Reason legislates a priori for freedom and its own causality, as the supersensible in the subject, for an unconditioned practical cognition. The domain of the concept of nature under the one legislation and that of the concept of freedom under the other are entirely barred from any mutual influence that they could have on each other by themselves (each in accordance with its fundamental laws) by the great chasm that separates the supersensible from the appearances" (5:195).

One way to understand the problem Kant is articulating here is to consider it once again in terms of the crisis of the Enlightenment.[27] The crisis was that modern science threatened to undermine traditional moral and religious beliefs, and Kant's response is to argue that in fact these essential interests of humanity are consistent with one another when reason is granted sovereignty and practical reason is given primacy over speculative reason. But the transcendental idealist framework within which Kant develops this response seems to purchase the consistency of these interests

at the price of sacrificing a unified view of the world and our place in it. If science applies only to appearances, while moral and religious beliefs refer to things in themselves or "the supersensible," then how can we integrate these into a single conception of the world that enables us to transition from the one domain to the other? Kant's solution is to introduce a third a priori cognitive faculty, which he calls the reflecting power of judgment, that gives us a teleological perspective on the world. Reflecting judgment provides the concept of teleology or purposiveness that bridges the chasm between nature and freedom, and thus unifies the theoretical and practical parts of Kant's philosophy into a single system (5:196–197).

It is important to Kant that a third faculty independent of both understanding and reason provides this mediating perspective, because he holds that we do not have adequate theoretical grounds for attributing objective teleology to nature itself, and yet regarding nature as teleological solely on moral grounds would only heighten the disconnect between our scientific and moral ways of viewing the world. Theoretical grounds do not justify us in attributing objective teleology to nature, because it is not a condition of self-consciousness that our understanding construct experience in accordance with the concept of teleology, which is not among Kant's categories or the principles of pure understanding that ground the fundamental laws of nature. That is why his theoretical philosophy licenses us only in attributing mechanical causation to nature itself. In this respect, Kant is sympathetic to the dominant strain in modern philosophy that banishes final causes from nature and instead treats nature as nothing but matter in motion, which can be fully described mathematically. But Kant wants somehow to reconcile this mechanistic view of nature with a conception of human agency that is essentially teleological. For as we saw in the previous section, Kant holds that every human action has an end and that the sum of all moral duties is to promote the highest good. It is essential to Kant's approach, however, to maintain the autonomy of both understanding (in nature) and reason (in morality),

without allowing either to encroach on the other's domain, and yet to harmonize them in a single system. This harmony can be orchestrated only from an independent standpoint, from which we do not judge how nature is constituted objectively (that is the job of understanding) or how the world ought to be (the job of reason), but from which we merely regulate or reflect on our cognition in a way that enables us to regard it as systematically unified. According to Kant, this is the task of reflecting judgment, whose a priori principle is to regard nature as purposive or teleological, "but only as a regulative principle of the faculty of cognition" (5:197).

7.2 The purposiveness of nature

In the Critique of the Power of Judgment, Kant discusses four main ways in which reflecting judgment leads us to regard nature as purposive: first, it leads us to regard nature as governed by a system of empirical laws; second, it enables us to make aesthetic judgments; third, it leads us to think of organisms as objectively purposive; and, fourth, it ultimately leads us to think about the final end of nature as a whole.[28]

First, reflecting judgment enables us to discover empirical laws of nature by leading us to regard nature as if it were the product of intelligent design (5:179–186). We do not need reflecting judgment to grasp the a priori laws of nature based on our categories, such as that every event has a cause. But in addition to these a priori laws nature is also governed by particular, empirical laws, such as that fire causes smoke, which we cannot know without consulting experience. To discover these laws, we must form hypotheses and devise experiments on the assumption that nature is governed by empirical laws that we can grasp (Bxiii–xiv). Reflecting judgment makes this assumption through its principle to regard nature as purposive for our understanding, which leads us to treat nature as if its empirical laws were designed to be understood by us (5:180–181). Since

this principle only regulates our cognition but is not constitutive of nature itself, this does not amount to assuming that nature really is the product of intelligent design, which according to Kant we are not justified in believing on theoretical grounds. Rather, it amounts only to approaching nature in the practice of science as if it were designed to be understood by us. We are justified in doing this because it enables us to discover empirical laws of nature. But it is only a regulative principle of reflecting judgment, not genuine theoretical knowledge, that nature is purposive in this way.

Second, Kant thinks that aesthetic judgments about both beauty and sublimity involve a kind of purposiveness, and that the beauty of nature in particular suggests to us that nature is hospitable to our ends. According to his aesthetic theory, we judge objects to be beautiful not because they gratify our desires, since aesthetic judgments are disinterested, but rather because apprehending their form stimulates what he calls the harmonious "free play" of our understanding and imagination, in which we take a distinctively aesthetic pleasure (5:204–207, 217–218, 287). So beauty is not a property of objects, but a relation between their form and the way our cognitive faculties work. Yet we make aesthetic judgments that claim intersubjective validity because we assume that there is a common sense that enables all human beings to communicate aesthetic feeling (5:237–240, 293–296). Beautiful art is intentionally created to stimulate this universally communicable aesthetic pleasure, although it is effective only when it seems unintentional (5:305–307). Natural beauty, however, is unintentional: landscapes do not know how to stimulate the free play of our cognitive faculties, and they do not have the goal of giving us aesthetic pleasure. In both cases, then, beautiful objects appear purposive to us because they give us aesthetic pleasure in the free play of our faculties, but they also do not appear purposive because they either do not or do not seem to do this intentionally. Kant calls this relation between our cognitive faculties and the formal qualities of objects that we judge to be beautiful

"subjective purposiveness" (5:221). Although it is only subjective, the purposiveness exhibited by natural beauty in particular may be interpreted as a sign that nature is hospitable to our moral interests (5:300). Moreover, Kant also interprets the experience of sublimity in nature as involving purposiveness. But in this case it is not so much the purposiveness of nature as our own purpose or "vocation" as moral beings that we become aware of in the experience of the sublime, in which the size and power of nature stand in vivid contrast to the superior power of our reason (5:257–260, 267–269).

Third, Kant argues that reflecting judgment enables us to regard living organisms as objectively purposive, but only as a regulative principle that compensates for our inability to understand them mechanistically, which reflects the limitations of our cognitive faculties rather than any intrinsic teleology in nature. We cannot understand organisms mechanistically because they are "self-organizing" beings, whose parts are "combined into a whole by being reciprocally the cause and effect of their form" (5:373–374). The parts of a watch are also possible only through their relation to the whole, but that is because the watch is designed and produced by some rational being. An organism, by contrast, produces and sustains itself, which is inexplicable to us unless we attribute to organisms purposes by analogy with human art (5:374–376). But Kant claims that it is only a regulative principle of reflecting judgment to regard organisms in this way, and that we are not justified in attributing objective purposiveness to organisms themselves, since it is only "because of the peculiar constitution of my cognitive faculties [that] I cannot judge about the possibility of those things and their generation except by thinking of a cause for these acts in accordance with intentions" (5:397–398). Specifically, we cannot understand how a whole can be the cause of its own parts because we depend on sensible intuition for the content of our thoughts and therefore must think the particular (intuition) first by subsuming it under the general (a concept). To see that this is just a limitation of the human, discursive

intellect, imagine a being with an intuitive understanding whose thought does not depend, as ours does, on receiving sensory information passively, but rather creates the content of its thought in the act of thinking it. Such a (divine) being could understand how a whole can be the cause of its parts, since it could grasp a whole immediately without first thinking particulars and then combining them into a whole (5:401–410). Therefore, since we have a discursive intellect and cannot know how things would appear to a being with an intuitive intellect, and yet we can only think of organisms teleologically, which excludes mechanism, Kant now says that we must think of both mechanism and teleology only as regulative principles that we need to explain nature, rather than as constitutive principles that describe how nature is intrinsically constituted (5:410ff.).

Fourth, Kant concludes the Critique of the Power of Judgment with a long appendix arguing that reflecting judgment supports morality by leading us to think about the final end of nature, which we can only understand in moral terms, and that conversely morality reinforces a teleological conception of nature. Once it is granted on theoretical grounds that we must understand certain parts of nature (organisms) teleologically, although only as a regulative principle of reflecting judgment, Kant says we may go further and regard the whole of nature as a teleological system (5:380–381). But we can regard the whole of nature as a teleological system only by employing the idea of God, again only regulatively, as its intelligent designer. This would be to attribute what Kant calls external purposiveness to nature – that is, to attribute purposes to God in creating nature (5:425). What, then, is God's final end in creating nature? According to Kant, the final end of nature must be human beings, but only as moral beings (5:435, 444–445). This is because only human beings use reason to set and pursue ends, using the rest of nature as means to their ends (5:426–427). Moreover, Kant claims that human happiness cannot be the final end of nature, because as we have seen he holds that happiness is not unconditionally valuable (5:430–431). Rather, human life has value

not because of what we passively enjoy, but only because of what we actively do (5:434). We can be fully active and autonomous, however, only by acting morally, which implies that God created the world so that human beings could exercise moral autonomy. Since we also need happiness, this too may be admitted as a conditioned and consequent end, so that reflecting judgment eventually leads us to the highest good (5:436). But reflection on conditions of the possibility of the highest good leads again to Kant's moral argument for belief in God's existence (he now omits immortality), which in turn reinforces the teleological perspective on nature with which reflecting judgment began.

Thus Kant argues that although theoretical and practical philosophy proceed from separate and irreducible starting points – self-consciousness as the highest principle for our cognition of nature, and the moral law as the basis for our knowledge of freedom – reflecting judgment unifies them into a single, teleological worldview that assigns preeminent value to human autonomy.

Bibliography

Primary Literature

Works by Kant

- The standard German edition of Kant's works is: Königlichen Preußischen (later Deutschen) Akademie der Wissenschaften (ed.), 1900–, Kants gesammelte Schriften, Berlin: Georg Reimer (later Walter De Gruyter).
- The best English edition of Kant's works is: P. Guyer and A. Wood (eds.), 1992–, *The Cambridge Edition of the Works of Immanuel Kant*, Cambridge: Cambridge University Press. Its individual volumes are:

Georg Wilhelm Friedrich Hegel

First published Thu Feb 13, 1997; substantive revision Tue Aug 4, 2015

Along with J.G. Fichte and, at least in his early work, F.W.J. von Schelling, Hegel (1770–1831) belongs to the period of German idealism in the decades following Kant. The most systematic of the post-Kantian idealists, Hegel attempted, throughout his published writings as well as in his lectures, to elaborate a comprehensive and systematic philosophy from a purportedly logical starting point. He is perhaps most well-known for his teleological account of history, an account that was later taken over by Marx and "inverted" into a materialist theory of an historical development culminating in communism. While *idealist philosophies* in Germany post-dated Hegel (Beiser 2014), the movement commonly known as German idealism effectively ended with Hegel's death. Certainly since the revolutions in logical thought from the turn of the twentieth century, the logical side of Hegel's thought has been largely forgotten, although his political and social philosophy and theological views have continued to find interest and support. Since the 1970s, however, a degree of more general philosophical interest in Hegel's systematic thought has been revived.

1. Life, Work, and Influence

Born in 1770 in Stuttgart, Hegel spent the years 1788–1793 as a student in nearby Tübingen, studying first philosophy, and then theology, and forming friendships with fellow students, the future great romantic poet Friedrich Hölderlin (1770–1843) and Friedrich von Schelling (1775–1854), who, like Hegel, would become one of the major figures of the

German philosophical scene in the first half of the nineteenth century. These friendships clearly had a major influence on Hegel's philosophical development, and for a while the intellectual lives of the three were closely intertwined.

After graduation Hegel worked as a tutor for families in Bern and then Frankfurt, where he was reunited with Hölderlin. Until around 1800, Hegel devoted himself to developing his ideas on religious and social themes, and seemed to have envisaged a future for himself as a type of modernising and reforming educator, in the image of figures of the German Enlightenment such as Lessing and Schiller. Around the turn of the century, however, under the influence of Hölderlin and Schelling, his interests turned more to issues arising from the critical philosophy initiated by Immanuel Kant (1724–1804) and developed by J.G. Fichte (1762–1814). In the 1790s the University of Jena had become a center for the development of critical philosophy due to the presence of K.L. Reinhold (1757–1823) and then Fichte, who taught there from 1794 until his dismissal on the grounds of atheism at the end of the decade. By that time, Schelling, who had first been attracted to Jena by the presence of Fichte, had become an established figure at the university. In 1801 Hegel moved to Jena to join Schelling, and in same year published his first philosophical work, *The Difference between Fichte's and Schelling's System of Philosophy*, in which he argued that Schelling had succeeded where Fichte had failed in the project of systematizing and thereby completing Kant's transcendental idealism. In 1802 and 1803 Hegel and Schelling worked closely together, editing the *Critical Journal of Philosophy*, and on the basis of this association Hegel came to be dogged for many years by the reputation of being a "mere" follower of Schelling (who was five years his junior).

By late 1806 Hegel had completed his first major work, the *Phenomenology of Spirit* (published 1807), which showed a divergence

from his earlier, seemingly more Schellingian, approach. Schelling, who had left Jena in 1803, interpreted a barbed criticism in the *Phenomenology*'s preface as aimed at him, and their friendship abruptly ended. The occupation of Jena by Napoleon's troops as Hegel was completing the manuscript restricted the activities of the university and Hegel departed. Now without a university appointment he worked for a short time, apparently very successfully, as an editor of a newspaper in Bamberg, and then from 1808–1815 as the headmaster and philosophy teacher at a gymnasium (high school) in Nuremberg. During his time at Nuremberg he married and started a family, and wrote and published his *Science of Logic*. In 1816 he managed to return to his university career by being appointed to a chair in philosophy at the University of Heidelberg, but shortly after, in 1818, he was offered and took up the chair of philosophy at the University of Berlin, the most prestigious position in the German philosophical world. In 1817, while in Heidelberg he published the *Encyclopaedia of the Philosophical Sciences*, a systematic work in which an abbreviated version of the earlier *Science of Logic* (the *Encyclopaedia Logic* or *Lesser Logic*) was followed by the application of its principles to the philosophy of nature and the philosophy of spirit. In 1821 in Berlin Hegel published his major work in political philosophy, *Elements of the Philosophy of Right*, based on lectures given at Heidelberg but ultimately grounded in the section of the *Encyclopaedia Philosophy of Spirit* dealing with *objective spirit*. During the following ten years up to his death in 1831 Hegel enjoyed celebrity at Berlin, and published subsequent versions of the *Encyclopaedia*. After his death versions of his lectures on philosophy of history, philosophy of religion, aesthetics, and the history of philosophy were published.

After Hegel's death, Schelling, whose reputation had long since been eclipsed by that of Hegel, was invited to take up the chair at Berlin, reputedly because the government of the day had wanted to counter the influence that Hegelian philosophy had exerted on a generation of

students. Since the early period of his collaboration with Hegel, Schelling had become more religious in his philosophising and criticised the rationalism of Hegel's philosophy. During this time of Schelling's tenure at Berlin, important forms of later critical reaction to Hegelian philosophy developed. Hegel himself had been a supporter of progressive but non-revolutionary politics, but his followers divided into factions broadly groupable as those of the left, right and centre (Toews 1985); from the *left*, Karl Marx was to develop his own purported scientific approach to society and history which appropriated many Hegelian ideas into a *materialistic* outlook. (Later, especially in reaction to orthodox Soviet versions of Marxism, many so-called Western Marxists re-incorporated further Hegelian elements back into their forms of Marxist philosophy.) Many of Schelling's own criticisms of Hegel's rationalism found their way into subsequent *existentialist* thought, especially *via* the writings of Kierkegaard, who had attended Schelling's lectures. Furthermore, the interpretation Schelling offered of Hegel during these years itself helped to shape subsequent generations' understanding of Hegel, contributing to the orthodox or traditional understanding of Hegel as a metaphysical thinker in the pre-Kantian dogmatic sense.

In academic philosophy, Hegelian idealism had seemed to collapse dramatically after 1848 and the failure of the revolutionary movements of that year, but underwent a revival in both Great Britain and the United States in the last decades of the nineteenth century. In Britain, where philosophers such as T.H. Green and F.H. Bradley had developed metaphysical ideas which they related back to Hegel's thought, Hegel came to be one of the main targets of attack by the founders of the emerging "analytic" movement, Bertrand Russell and G.E. Moore. For Russell, the revolutionary innovations in logic starting in the last decades of the nineteenth century had destroyed Hegel's metaphysics by overturning the Aristotelian logic on which, so Russell claimed, it was based, and in line with this dismissal, Hegel came to be seen within the

analytic movement as an historical figure of little genuine philosophical interest. To some degree, analogous things could be said of Hegel's reception from within the twentieth-century phenomenological tradition that developed in continental Europe, but although marginalized within such core areas of mainstream academic philosophy, Hegel nevertheless continued to be a figure of interest within other philosophical movements such as existentialism and Marxism. In France, a version of Hegelianism came to influence a generation of thinkers, including Jean Hyppolite, Jean-Paul Sartre and the psychoanalyst, Jacques Lacan, largely through the lectures of Alexandre Kojève. However, a later generation of French philosophers coming to prominence in the 1960s tended to react against Hegel in ways analogous to those in which early analytic philosophers had reacted against the Hegel who had influenced their predecessors. In Germany, having lapsed in the second half of the nineteenth century, interest in Hegel was revived at the turn of the twentieth with the historical work of Wilhelm Dilthey, and important Hegelian elements were incorporated within the approaches of thinkers of the Frankfurt School, such as Theodor Adorno, and later, Jürgen Habermas, as well as within the Heidegger-influenced *hermeneutic* approach of H.-G. Gadamer. In Hungary, similar Hegelian themes were developed by Georg Lukács and later thinkers of the Budapest School. In the 1960s the German philosopher Klaus Hartmann developed what was termed a *non-metaphysical* interpretation of Hegel which, together with the work of Dieter Henrich and others, played an important role in the revival of interest in Hegel in academic philosophy in the second half of the century. Within English-speaking philosophy, the final quarter of the twentieth century saw something of a revival of serious interest in Hegel's philosophy with important works appearing such as those by H.S. Harris, Charles Taylor, Robert Pippin and Terry Pinkard in North America, and Stephen Houlgate and Robert Stern in Great Britain. By the close of the twentieth century, even within core logico-metaphysical areas of analytic

philosophy, a number of individuals such as Robert Brandom and John McDowell had started to take Hegel seriously as a significant modern philosopher, although generally within analytic circles a favorable reassessment of Hegel has still a long way to go.

2. Hegel's Philosophy

Hegel's own pithy account of the nature of philosophy given in the *Preface* to his *Elements of the Philosophy of Right* captures a characteristic tension in his philosophical approach and, in particular, in his approach to the nature and limits of human cognition. "Philosophy", he says there, "is its own time comprehended in thoughts" (PR: 21).

On the one hand we can clearly see in the phrase "its own time" the suggestion of an historical or cultural conditionedness and variability which applies even to the highest form of human cognition, philosophy itself. The contents of philosophical knowledge, we might suspect, will come from the historically changing contents of its cultural context. On the other, there is the hint of such contents being raised to some higher level, presumably higher than other levels of cognitive functioning such as those based in everyday perceptual experience, for example, or those characteristic of other areas of culture such as art and religion. This higher level takes the form of conceptually articulated *thought*, a type of cognition commonly taken as capable of having purportedly eternal contents (think of Plato and Frege, for example). In line with such a conception, Hegel sometimes referred to the task of philosophy as that of recognising *the concept* (*Der Begriff*) in the mere representations (*Vorstellungen*) of everyday life.

This antithetical combination within human cognition of the temporally-conditioned and the eternal, a combination which reflects a broader conception of the human being as what Hegel describes elsewhere as a

"finite-infinite", (SL: 114) has led to Hegel being regarded in different ways by different types of philosophical readers. For example, an historically-minded pragmatist like Richard Rorty, distrustful of all claims or aspirations to the so-called God's-eye view, could praise Hegel as a philosopher who had introduced this historically reflective dimension into philosophy (and set it on the characteristically romantic path which has predominated in modern continental philosophy) but who had unfortunately still remained bogged down in the remnants of the Platonistic idea of the search for ahistorical truths (Rorty 1982). Those adopting such an approach to Hegel tend to have in mind the (relatively) young author of the *Phenomenology of Spirit* and have tended to dismiss as "metaphysical" later and more systematic works like the *Science of Logic*. In contrast, the British Hegelian movement at the end of the nineteenth century tended to ignore the *Phenomenology* and the more historicist dimensions of his thought, and found in Hegel a systematic metaphysician whose *Logic* provided the basis for a definitive philosophical ontology. This latter traditional *metaphysical* view of Hegel dominated Hegel reception for most of the twentieth century, but from the 1980s came to be challenged by scholars who offered an alternative non-metaphysical, post-Kantian view. By "non-metaphysical" these thinkers had in mind metaphysics in the sense that Kant had been critical of, a point sometimes missed by critics. But in turn, this post-Kantian reading has been challenged by a *revised metaphysical* view, critical of the purported *over*-assimilation of Hegel to Kant by the post-Kantians. In the revised metaphysical view, appeal is often made to Aristotelian or Spinozist conceptual realist features of Hegel's thought, as well as to features of recent analytic metaphysics.

Before surveying these competing views, however, something needs to be said about the confusing term "idealism", and about the variety of idealism that is characteristic of Hegel and other German idealists.

2.1 Background: Idealism as understood in the German tradition

"Idealism" is a term that had been used sporadically by Leibniz and his followers to refer to a type of philosophy that was opposed to *materialism*. Thus, for example, Leibniz had contrasted *Plato* as an idealist with *Epicurus* as a materialist. The opposition to materialism here, together with the fact that in the English-speaking world the Irish philosopher and clergyman George Berkeley (1685–1753) is often taken as a prototypical idealist, has given rise to the assumption that idealism is necessarily an immaterialist doctrine. This assumption, however, is mistaken. With the possible exception of Leibniz, the idealism of the Germans was not committed to the type of doctrine found in Berkeley according to which immaterial minds, both infinite (God's) and finite (those of humans), were the ultimately real entities, with *apparently* material things to be understood as reducible to *states* of such minds—that is, to ideas in the sense meant by the British empiricists.

As Leibniz's use of Plato to exemplify idealism suggests, idealists in the German tradition tended to hold to the reality or objectivity of ideas in the Platonic sense, and for Plato, it would seem, such ideas were not conceived as *in* any mind at all—not even the mind of Plato's *god*. The type of picture found in Berkeley was only to be found in certain late antique Platonists and, especially, early Christian Platonists like Saint Augustine, Bishop of Hippo. But especially for the German idealists like Hegel, Plato's philosophy was understood through the lenses of more Aristotelian varieties of neo-Platonism, which pictured the thoughts of a divine mind as immanent in matter, and not as contained in some purely immaterial or spiritual mind. It thus had features closer to the more pantheistic picture of divine thought found in Spinoza, for example, for whom matter and mind were attributes of the one substance.

Even for Leibniz, whose later monadological metaphysics was perhaps closer to Berkeley's immaterialist philosophy, an opposition to materialism didn't necessarily imply immaterialism. Leibniz had resisted Descartes' postulation of distinct spiritual and material substances, treating corporeal bodies as inseparable combinations of form and matter after the manner of Aristotle. The materialists to which he was opposed (mechanistic corpuscularists of his time) conceived of unformed matter as a type of self-subsistent substance, and it seems to have been *that* conception to which he was opposed, at least in some periods of his work, not the reality of matter *per se*. Leibniz's combination of Platonic and Aristotelian notions played a role in the thought of the later idealists, giving their opposition to materialism its distinctive character. These anti-immaterialist features of the idealism of the Germans became more prominent in the post-Kantian period as the moved progressively away from the more subjectivistic features of Leibniz's thought (Beiser 2002).

2.2 The *traditional* metaphysical view of Hegel's philosophy

Given the understanding of Hegel that predominated at the time of the birth of analytic philosophy, together with the fact that early analytic philosophers were rebelling precisely against Hegelianism so understood, the interpretation of Hegel encountered in discussions within analytic philosophy is often that of the late nineteenth-century interpretation. In this picture, Hegel is seen as offering a metaphysico-religious view of God *qua* Absolute Spirit, as the ultimate reality that we can come to know through pure thought processes alone. In short, Hegel's philosophy is treated as exemplifying the type of pre-critical or *dogmatic* metaphysics against which Kant had reacted in his *Critique of Pure Reason*, and as a return to a more religiously driven conception of philosophy to which Kant had been opposed.

There is much that can be found in Hegel's writings that seems to support this view. In his lectures during his Berlin period one comes across claims such as the one that philosophy "has no other object but God and so is essentially rational theology" (Aes I: 101). Indeed, Hegel often seems to invoke imagery consistent with the types of neo-Platonic conceptions of the universe that had been common within Christian mysticism, especially in the German states, in the early modern period. The peculiarity of Hegel's form of idealism, on this account, lies in his idea that the mind of God becomes actual only via its particularization in the minds of "his" finite material creatures. Thus, in our consciousness of God, we somehow serve to realize his *own* self-consciousness, and, thereby, his own perfection. In English-language interpretations, such a picture is effectively found in the work of Charles Taylor (1975) and Michael Rosen (1984), for example. With its dark mystical roots, and its overtly religious content, it is hardly surprising that the philosophy of Hegel so understood has rarely been regarded as a live option within the largely secular and scientific conceptions of philosophy that have been dominant in the twentieth century.

An important consequence of Hegel's metaphysics, so understood, concerns history and the idea of historical development or progress, and it is as an advocate of an idea concerning the logically-necessitated teleological course of history that Hegel is most often derided. To critics, such as Karl Popper in his popular post-war *The Open Society and its Enemies* (1945), Hegel had not only advocated a disastrous political conception of the state and the relation of its citizens to it, a conception prefiguring twentieth-century totalitarianism, but he had also tried to underpin such advocacy with dubious theo-logico-metaphysical speculations. With his idea of the development of *spirit* in history, Hegel is seen as literalising a way of talking about different cultures in terms of their spirits, of constructing a developmental sequence of epochs typical of nineteenth-century ideas of linear historical progress, and then enveloping

this story of human progress in terms of one about the developing self-conscious of the cosmos-God itself.

As the bottom line of such an account concerned the evolution of states of a mind (God's), such an account is clearly an idealist one, but not in the sense, say, of Berkeley. The pantheistic legacy inherited by Hegel meant that he had no problem in considering an objective *outer* world beyond any particular subjective mind. But this objective world itself had to be understood as conceptually informed: it was *objectified* spirit. Thus in contrast to Berkeleian *subjective* idealism it became common to talk of Hegel as incorporating the *objective* idealism of views, especially common among German historians, in which social life and thought were understood in terms of the conceptual or spiritual structures that informed them. But in contrast to both forms of idealism, Hegel, according to this reading, postulated a form of *absolute* idealism by including both subjective life and the objective cultural practices on which subjective life depended within the dynamics of the development of the self-consciousness and self-actualisation of God, the *Absolute* Spirit.

Despite this seemingly dominant theological theme, Hegel was still seen by many as an important precursor of other more characteristically secular strands of modern thought such as existentialism and Marxist materialism. Existentialists were thought of as taking the idea of the finitude and historical and cultural dependence of individual subjects from Hegel, and as leaving out all pretensions to the Absolute, while Marxists were thought of as taking the historical dynamics of the Hegelian picture but reinterpreting this in materialist rather than idealist categories. As for understanding Hegel himself, the traditional metaphysical view remained the dominant interpretative approach of Hegel scholars throughout much of the twentieth century. In the last quarter of the century, however, it came to be vigorously questioned, with a variety of interpreters putting forward very different accounts of the basic nature of Hegel's

philosophical project. While a number of interpretations of Hegel have emerged during this period in an effort to acquit him of implausible metaphysico-theological views, one prominent tendency has been to stress the continuity of his ideas with the "critical philosophy" of Immanuel Kant.

2.3 The post-Kantian (sometimes called the non-metaphysical) view of Hegel

Least controversially, it is often claimed that either particular *works*, such as the *Phenomenology of Spirit*, or particular *areas* of Hegel's philosophy, especially his ethical and political philosophy, can be understood as standing independently of the type of unacceptable metaphysical system sketched above. Thus it is commonly asserted that implicit within the metaphysical Hegel is an anti-metaphysical philosopher struggling to get out—one potentially capable of beating the critical Kant at his own game.

More controversially, one now finds it argued that the traditional picture is simply wrong at a more general level, and that Hegel, even in his systematic thought, was not committed to the bizarre, teleological *spirit monism* that has been traditionally attributed to him because he was free of the type of traditional metaphysical commitments that had been criticized by Kant. Prominent among such interpretations has been the so-called post-Kantian interpretation advanced by North American Hegel scholars Robert Pippin (1989, 2008, 2010) and Terry Pinkard (1994, 2000, 2012). From an explicitly analytic perspective, broadly similar views have been put forward by Robert Brandom (2002, 2007, 2014) and John McDowell (2006). Thus while the traditional view sees Hegel as exemplifying the very type of metaphysical speculation that Kant successfully criticised, the post-Kantian view regards him as both accepting and *extending* Kant's critique, ultimately turning it against the residual dogmatically metaphysical aspects of Kant's own philosophy.

In Hegel, the non-traditionalists argue, one can see the ambition to bring together the universalist dimensions of Kant's transcendental program with the culturally contextualist conceptions of his more historically and relativistically-minded contemporaries, resulting in his controversial conception of spirit, as developed in his *Phenomenology of Spirit*. With this notion, it is claimed, Hegel was essentially attempting to answer the Kantian question of the conditions of rational human mindedness, rather than being concerned with giving an account of the developing self-consciousness of God. But while Kant had limited such conditions to formal abstractly conceived structures of the mind, Hegel extended them to include aspects of historically and socially determined forms of embodied human existence.

2.4 The *revised* metaphysical view of Hegel

Not surprisingly, the strong post-Kantian interpretation of Hegel has been resisted by defenders of the more traditional approach, who have argued against the plausibility of attempting to rehabilitate Hegel's philosophy by divesting it of any purportedly unacceptable metaphysical claims (see, for example, Beiser 2005 and Horstmann 2006). Proponents of the post-Kantian view, it is commonly said, are guilty of projecting onto Hegel views they would *like* to find there rather than what is actually to be found. However, the strong post-Kantian interpretation has *also* been challenged by a somewhat different version of the metaphysical reading by interpreters who, while recognizing the influence of Kant's critical philosophy of Hegel, emphasize Hegel's critique of Kant and affirm the irreducible role played by a form of metaphysics in Hegel's philosophy. Nevertheless, they share the post-Kantians' attempts to separate Hegel's views from the extravagant views traditionally ascribed to him and generally argue for the broad acceptability of Hegel's views from the perspective of the present. Here one tends to find interpreters attributing to Hegel some type of conceptual realism, sometimes appealing to

contemporary *analytic* metaphysics for the legitimacy of metaphysics conceived as inquiry into the fundamental features or structures of the world itself. Among the interpreters advancing something like this *revised* metaphysical view might be counted Stephen Houlgate (2005b), Robert Stern (2002, 2009), Kenneth Westphal (2003), James Kreines (2006, 2008) and Christopher Yeomans (2012).

On a number of points, the proponents of the revised conceptual realist metaphysical interpretation will agree with advocates of the post-Kantian non-metaphysical approach. First, they tend to agree in dismissing much of the extravagant metaphysics traditionally ascribed to Hegel. Generally they don't find in Hegel the type of classical teleological spirit monism central to, say, Taylor's interpretation. Next, they stress the importance for Hegel of Kant's critique of metaphysics. Both think that Hegel took Kant's critique seriously, and in turn subjected that critique itself to a telling *meta-critique*, showing that Kant himself was not free from the sorts of ungrounded metaphysical assumptions he criticized in others. However, while the post-Kantians interpret Hegel's criticisms of Kant as suggesting that Hegel thereby realized or completed Kant's critical intention, creating a form of philosophizing purged of metaphysics, proponents of the revised metaphysical interpretation typically see his criticism of Kant as involving a rejection of Kant's anti-metaphysical attitude, and as reestablishing, on a new basis, a metaphysical program originally derived from Aristotle (e.g., Stern) or Spinoza (e.g., Houlgate).

While it is for the most part clear what sets both post-Kantians and conceptual realists against the traditional view, it is still not clear which issues dividing them are substantive and which are ultimately verbal. After all, Kant himself was not critical of metaphysics *per se*. His claim was that existing (so-called dogmatic) metaphysics was in a state analogous to that in which, say, physics had been in before the scientific revolution of sixteenth and seventeenth centuries. Rather than wanting to eliminate

metaphysics, after the style, say, of Hume or the modern logical positivists, Kant had wanted to put metaphysics itself on a secure scientific basis analogous to what Galileo and Newton had achieved for physics. Thus the very idea of an *Hegelian metaphysics* is in no way straightforwardly incompatible with the project of a post-Kantian completion of Kant's critical program. The relevant differences between revised metaphysical and the non-metaphysical views would need to be established with respect to such particular issues as, for example, the nature of acceptably Kantian metaphysical claims.

3. Hegel's Published Works

We may think of there being five different types of work that make up Hegel's published corpus. First, there are Hegel's two major stand-alone books written for publication already mentioned—*Phenomenology of Spirit* (1807) and *Science of Logic* (1812–18). In the next category are works that were published at the time as *handbooks* for use in student teaching such as the *Encyclopaedia of Philosophical Sciences* first published in 1817 while he was teaching at Heidelberg and subsequently revised and republished in 1827 and again in 1830, and *Elements of the Philosophy of Right*, effectively an expansion of a section of the *Encyclopaedia* and published in 1820 after his move to Berlin. (Transcripts of his earlier lectures on this topic delivered in Heidelberg have also since been published.) Along with the *Encyclopaedia* and the *Philosophy of Right* might be added similar teaching-related writings from the Jena period, prepared as lectures but only published as such much later. The third major category is formed by posthumously published lecture courses from his time at the University of Berlin, which, after Hegel's death, were assembled by editors from his lecture notes and from student transcripts of the lectures as delivered—these include his lectures on the *Philosophy* of *Nature*, *Philosophy of Spirit*, *Philosophy of History*,

Aesthetics, *Philosophy of Religion*, and *History of Philosophy*. Next might be considered various miscellaneous essays and short works published during his career, and finally we can count Hegel's early works, written in the period between his student years at Tübingen and his move to Jena, and predominantly on religious and political themes (ETW). Here we will restrict the discussion to the first three categories.

3.1 Books

3.1.1 *Phenomenology of Spirit*

The term "phenomenology" had been coined by the Swiss mathematician (and Kant correspondent) J.H. Lambert (1728–1777) in his *The New Organon* of 1764, and in a letter to Lambert, sent to accompany a copy of his Inaugural Dissertation (1770), Kant had proposed his own project of a "general phenomenology" as a necessary propaedeutic presupposed by the science of metaphysics. Such a phenomenology was meant to determine the "validity and limitations" of what he called the "principles of sensibility", principles he had (he thought) shown in the accompanying work to be importantly different to those of conceptual thought. The term clearly suited Kant as he had distinguished the phenomena known through the faculty of sensibility from the noumena known purely conceptually. This envisioned phenomenology seems to coincide roughly with what he was to eventually describe as a *critique of pure reason*, although Kant's thought had gone through important changes by the time that he came to publish the work of that name (1781, second edition 1787). Perhaps because of this he never again used the term "phenomenology" for quite this purpose.

There is clearly some continuity between Kant's notion and Hegel's project. In a sense Hegel's phenomenology is a study of *phenomena* (although this is not a realm he would contrast with that of noumena) and

Hegel's *Phenomenology of Spirit* is likewise to be regarded as a type of propaedeutic to philosophy rather than an exercise *in* or work *of* philosophy. It is meant to function as an induction or education of the reader to the *standpoint* of purely conceptual thought from which philosophy can be done. As such, its structure has been compared to that of a *Bildungsroman* (educational novel), having an abstractly conceived protagonist—the bearer of an evolving series of so-called shapes of consciousness or the inhabitant of a series of successive phenomenal worlds—whose progress and set-backs the reader follows and learns from. Or at least this is how the work sets out: in the later sections the earlier series of shapes of consciousness becomes replaced with what seem more like configurations of human *social life*, and the work comes to look more like an account of interlinked forms of social existence and thought within which participants in such forms of social life conceive of themselves and the world. Hegel constructs a series of such shapes that maps onto the history of western European civilization from the Greeks to his own time.

The fact that this progression ends in the attainment of what Hegel refers to as Absolute Knowing, the standpoint from which real philosophy gets done, seems to support the traditionalist reading in which a triumphalist narrative of the growth of western civilization is combined with the theological interpretation of God's self-manifestation and self-comprehension. When Kant had broached the idea of a phenomenological propaedeutic to Lambert, he himself had still believed in the project of a purely conceptual metaphysics achievable by the use of the regressive or *analytic* method, but this project conceived as an exercise in theoretical reason was just what Kant in his later critical philosophy had come to disavow. Traditional readers of Hegel thus see the *Phenomenology*'s *telos* as attesting to Hegel's pre-Kantian (that is, pre-critical) outlook, and his embrace of the metaphysical project that Kant famously came to dismiss as illusory. Supporters of the *post-Kantian* interpretation of Hegel obviously interpret this work and its *telos* differently. For example, it has

been argued (e.g., Pinkard 1994) that what this history tracks is the development of a type of social existence which enables a unique form of rationality, in that in such a society all dogmatic bases of thought have been gradually replaced by a system in which all claims become open to rational self-correction, by becoming exposed to demands for conceptually-articulated justifications. As Pinkard had pointed out in that work, this was a conception of the normatively structured practices of human reason found in the American pragmatist Wilfrid Sellars, the inspiration behind the Hegelian dimensions of analytic philosophers such as Willem deVries (1988), Robert Brandom and John McDowell.

Something of Hegel's phenomenological method may be conveyed by the first few chapters, which are perhaps among its more conventionally philosophical parts (Westphal 2009). Chapters 1 to 3 effectively follow a developmental series of distinct shapes of consciousness—jointly epistemological and ontological attitudes articulated by criteria which are, regarded from one direction, criteria for certain *knowledge*, and from the other, criteria for the nature of the *objects* of such knowledge. In chapter 1, the attitude of *Sense-certainty* takes immediately given perceptual simples —the sort of role played by the so-called sense-data of early twentieth-century analytic epistemology, for example, with which a subject is purportedly acquainted as *bare thises*—as the fundamental objects known. By following this form of consciousness's attempts to make these implicit criteria *explicit*, we are meant to appreciate that any such contents, even the apparently most immediate ones, are in fact grasped *conceptually*, and so, in Hegel's terminology, their reception is actually mediated by the concepts with which they are grasped. Hegel is clear that these contents are not merely qualitative simples that are immediately *apprehended*, but *comprehended* instances of the conceptual determination of singularity [*Einzelheit*] (Phen: §91). Such a simple *this*, then, can also be understood as an instance of what the Medievals discussed as *thisness*—a general *property* of an individual thing's being identical to itself. One might

compare Hegel's point here to that expressed by Kant in his well known claim from *Critique of Pure Reason* (A51/B75), that *without* general concepts, *intuitions* (singular [*einzeln*] purportedly immediate mental representations), are *blind*. However, Hegel seems to want to make this point *without* relying on Kant's formal distinction *between* concepts and intuitions as different species of representation. The idea seems to be that for Hegel, *the same content* can play the roles played by both concepts and intuitions in Kant. (The lessons of this chapter have sometimes been likened to those of Wilfrid Sellars's famous criticism of the empiricist *myth of the given*.)

By the end of this chapter our protagonist consciousness (and by implication, we the audience to this drama) has learnt that the nature of consciousness cannot be as originally thought: rather than being immediate and singular, its contents must have some implicit universal (conceptual) aspect to them. The general *truth* that was learned about the apparent qualitative simples in Sense-certainty (that they were instances of generals) is now explicitly taken as the *truth of the object* of *Perception* (*Wahrnehmung*—in German this term having the connotations of taking (*nehmen*) to be true (*wahr*)). In contrast to the purported single object of Sense-certainty the object of Perception is taken *as* instantiating general properties: it is "a thing with many properties" (Phen: §112). But this can be conceived in a variety of ways: first, as a simple *bundle* of indifferent qualities (a picture associated with Plato), or as an underlying substrate *in which* these qualities somehow inhere (a picture associated with Aristotle). Predictably, problems will be revealed in these various different ways of thinking of the nature of those everyday objects of our experience.

As in the case of Sense-certainty, here in the case of Perception, by following the protagonist consciousness's efforts to make this implicit criterion explicit, we see how the criterion generates contradictions that eventually undermine it *as* a criterion for certainty. In fact, such collapse

into a type of self-generated skepticism is typical of all the shapes we follow in the work, and there seems something inherently skeptical about such reflexive cognitive processes. But this is not the type of skepticism that is typical of early modern philosophy, such as that used by Descartes in his attempt to find some foundation of *indubitability* on which genuine knowledge can be built (Forster 1989). As is clear from his treatment of ancient philosophy in the *Lectures on the History of Philosophy*, Hegel was attracted to the type of *dialectic* employed by Socrates in his efforts to get his interlocutors thinking about something beyond that given immediately in sensation (LHP II: 51), and implicit in the ancient form of skepticism that had been employed after Socrates (LHP II: 344). For Hegel, the ancient skeptics captured the skeptical moment of thought that is the means by which thought progresses *beyond* the particular categories that have given rise to contradictions. Just as in the way a new shape of thought, Perception, had been generated from the internal contradictions that emerged within Sense-certainty, the collapse of any given attitude will be accompanied by the emergence of some new implicit criterion that will be the basis of a new emergent attitude. In the case of Perception, the emergent new shape of consciousness, the Understanding, explored in Chapter 3, is a shape identified with the type of scientific cognition that, rather than remaining on the level of the perceived object, posits underlying *forces* involved in the production of the perceptual episode.

The transition from Chapter 3 to Chapter 4, The Truth of Self-Certainty, also marks a more general transition from Consciousness to *Self*-consciousness. It is in the course of Chapter 4 that we find what is perhaps the most well-known part of the *Phenomenology*, the account of the *struggle of recognition* in which Hegel examines the inter-subjective conditions which he sees as necessary for any form of consciousness. This is a topic that had first been taken up by Alexandre Kojève (1969), and which has been appealed to in non-Kojèvean ways recently by a number of non-traditional interpreters in order to give a quite different accounts of

Hegel's notion of *Spirit* (Honneth 1995; Ikäheimo and Laitinen 2011; Pippin 2008; Redding 1996; Williams 1992, 1997).

Like Kant, Hegel thinks that one's capacity to be *conscious* of some external object as something distinct from oneself requires the reflexivity of *self*-consciousness, that is, it requires one's awareness *of oneself* as a subject *for whom* something distinct, the object, is presented *as known* (a result emerging in Chapter 3). Hegel goes beyond Kant, however, and expanding on an idea found in Fichte, makes this requirement dependent on one's recognition (or acknowledgment—*Anerkennung*) *of other* self-conscious subjects *as* self-conscious subjects *for whom* any object of consciousness will be thought as *also* existing. One's self-consciousness, in fact, will be dependent on one's recognition of those others as similarly recognizing *oneself* as a self-conscious subject. Such complex patterns of mutual recognition constituting *objective spirit* thereby provide the social matrix within which individual self-consciousnesses can exist as such. It is in this way that the *Phenomenology* can change course, the earlier tracking of shapes of individual consciousness and self-consciousness effectively coming to be replaced by the tracking of distinct patterns of mutual recognition between subjects—shapes of spirit—that forms the ground for the existence of those individual consciousnesses/self-consciousnesses.

It is thus that Hegel has effected the transition from a phenomenology of the individual's subjective mind to one of *objective* spirit, thought of as culturally distinct objective patterns of social interaction to be analysed in terms of the patterns of reciprocal recognition they embody. ("Geist" can be translated as either "mind" or "spirit", but the latter, allowing a more cultural sense, as in the phrase "spirit of the age" ("Zeitgeist"), seems a more suitable rendering for the title.) But this is only worked out in the text gradually. We—the reading or so-called phenomenological we—can see how particular shapes of self-consciousness, such as that of the other-worldly religious self-consciousness (Unhappy Consciousness) with which

Chapter 4 ends, depend on certain institutionalised forms of mutual recognition, in this case one involving a priest who mediates between the self-conscious subject and that subject's God. But we are seeing this "from the outside", as it were: we still have to learn how actual self-consciousnesses could learn this of themselves. So we have to see how the protagonist self-consciousness could achieve this insight. It is to this end that we further trace the learning path of self-consciousness through the processes of *reason* (in Chapter 5) before objective spirit can become the explicit subject matter of Chapter 6 (Spirit).

Hegel's discussion of spirit starts from what he calls *Sittlichkeit* (translated as "ethical order" or "ethical substance"—"*Sittlichkeit*" being a nominalization from the adjectival (or adverbial) form "*sittlich*", "customary", from the stem "*Sitte*", "custom" or "convention".) Thus Hegel might be seen as adopting the viewpoint that since social life is ordered by customs we can approach the lives of those living in it in terms of the patterns of those customs or conventions themselves—the conventional practices, as it were, constituting specific, shareable *forms of life* made actual in the lives of particular individuals who had in turn *internalized* such general patterns in the process of acculturation. It is not surprising then that his account of spirit here starts with a discussion of religious and civic law. Undoubtedly it is Hegel's tendency to nominalise such abstract concepts in his attempt to capture the *concrete* nature of such patterns of conventional life, together with the tendency to then *personify* them (as in talking about spirit *becoming* self-conscious) that lends plausibility to the traditionalist understanding of Hegel. But for non-traditionalists it is not obvious that Hegel, in employing such phrases, is in any way committed to any metaphysical supra-individual conscious being or beings. To take an example, in the second section of the chapter on spirit, Hegel discusses human culture as the "world of self-alienated spirit". The idea seems to be that humans in society not only interact, but that they collectively create relatively enduring cultural products

(repeatable stories, stageable dramas, and so forth) within which members of that society can recognise patterns of their own communal life as so reflected. We might find intelligible the metaphor that such products "hold up a mirror to society" within which "the society can regard itself", without thinking we are thereby committed to some supra-individual unitary mind achieving self-consciousness. Furthermore, such cultural products themselves provide conditions allowing individuals to adopt particular cognitive attitudes by appropriating their resources. Thus, for example, the capacity to adopt the type of objective viewpoint demanded by Kantian morality (discussed in the final section of Spirit)—the capacity to see things, as it were, from a detached or universal point of view—might be enabled by engaging with spirit's "alienations" such as the myths and rituals of a religion professing a universal scope.

We might think that if Kant had written Hegel's *Phenomenology* he would have ended it at Chapter 6 with the modern moral subject as the *telos* of the story. For Kant, the practical knowledge of morality, orienting one within the noumenal world, exceeds the scope of theoretical knowledge, which had been limited to phenomena. Hegel, however, thought that philosophy had to unify theoretical and practical knowledge, and so the *Phenomenology* has further to go. Again, this is seen differently by traditionalists and revisionists. For traditionalists, Chapters 7, Religion and 8, Absolute Knowing, testify to Hegel's disregard for Kant's critical limitation of theoretical knowledge to empirical experience. Revisionists, on the other hand, tend to see Hegel as furthering the Kantian critique into the very coherence of a conception of an *in-itself* reality that is beyond the limits of our theoretical (but not practical) cognition. Rather than understand absolute knowing as the achievement of some ultimate God's-eye view of everything, the philosophical analogue to the connection with God sought in religion, post-Kantian revisionists see it as the accession to a mode of self-critical thought that has finally abandoned all non-questionable mythical *givens*, and which will only countenance reason-

giving argument as justification. However we understand this, absolute knowing is the standpoint to which Hegel has hoped to bring the reader in this complex work. This is the standpoint of *science*, the standpoint from which philosophy proper commences, and it commences in Hegel's next book, the *Science of Logic*.

3.1.2 *Science of Logic*

Hegel's *Science of Logic* is divided into three books, dealing with the topics *of being*, *essence*, and *the concept*, which appeared in 1812, 1813, and 1816 respectively. For most of the 20th century it was not received with the enthusiasm that often marked the reception of *Phenomenology of Spirit*. First, as a work of *logic* most have regarded it as radically outdated and relying on an Aristotelian approach that was definitively surpassed in the later nineteenth century—a view promoted especially by Bertrand Russell in the early years of the twentieth. Thus many readers sympathetic to particular doctrines in Hegel have attempted, contrary to Hegel's insistence, to quarantine his philosophical approach to particular areas from it. Recently, this skepticism has started to change.

Some advocate that the *Science of Logic* be read as a first-order *ontological* doctrine (Doz 1987) or as a category theory that *simultaneously* represents structures of being and thought (Houlgate 2005b), and so as having very little to do with what has traditionally been known as logic. Others argue that in contrast to the project of formal (or general) logic, it is best understood as a version of what Kant had called "transcendental logic" (di Giovanni 2010). In this sense it should thereby be thought of as a successor to Kant's "transcendental deduction of the categories" in the *Critique of Pure Reason* in which Kant attempted to derive a list of those non-empirical concepts, the *categories*, which he believed to be presupposed by all empirical judgments made by finite, discursive knowers like ourselves. In short, taking the logic as a *category*

theory opens up two general lines of interpretation: should the categories be understood as primarily *ontological* categories, as found in Aristotle, or as primarily categories revealing the necessary structure of *thought*, as in Kant? Those, such as the advocates of the revised metaphysical interpretation, interpreting Hegel as basically a metaphysician, typically stress the former, while post-Kantian interpreters typically stress the latter.

A glance at the table of contents of *Science of Logic* reveals the same triadic structuring among the categories or thought determinations discussed that has been noted among the shapes of consciousness in the *Phenomenology*. At the highest level of its branching structure there are the three books devoted to the doctrines of being, essence, and concept, while in turn, each book has three sections, each section containing three chapters, and so on. In general, each of these individual nodes deals with some particular category. In fact, Hegel's categorial triads appear to repeat Kant's *own* triadic way of articulating the categories in the Table of Categories (*Critique of Pure Reason* A80/B106) in which the third term in the triad in some way integrates the first two. (In Hegel's terminology, he would say that the first two were *sublated* [*aufgehoben*] in the third— while the first two are *negated* by the third, they continue to work within the context defined by it.) Hegel's later treatment of *the syllogism* found in Book 3, in which he follows Aristotle's own three-termed schematism of the syllogistic structure, repeats this triadic structure as does his ultimate analysis of its component concepts as the moments of universality, particularity, and singularity.

Reading into the first chapter of Book 1, Being, it is quickly seen that the *transitions* of the *Logic* broadly repeat those of the first chapters of the *Phenomenology*, now, however, as between the categories themselves rather than between conceptions of the respective *objects* of conscious experience. Thus, being is the thought determination with which the work commences because it *at first* seems to be the most immediate,

fundamental determination that characterises any possible thought content at all. (In contrast, *being* in the *Phenomenology*'s *Sense-certainty* chapter was described as the known *truth* of the purported immediate sensory given—the category that it was discovered to instantiate.) Whatever thought is about, that topic must in some sense *exist*. Like those purported simple *sensory givens* with which the *Phenomenology* starts, the category *being* looks to have no internal structure or constituents, but again in a parallel to the *Phenomenology*, it is the effort of thought to make this category explicit that both undermines it and brings about new ones. *Being* seems to be *both* immediate and simple, but it will show itself to be, in fact, *only* something in opposition to something else, *nothing*. The point seems to be that while the categories being and nothing seem both absolutely distinct and opposed, on reflection (and following Leibniz's principle of the identity of indiscernibles) they appear identical as no criterion can be invoked which differentiates them. The only way out of this paradox is to posit a third category within which they can coexist as *negated* (*Aufgehoben*) moments. This category is *becoming*, which saves thinking from paralysis because it accommodates both concepts. Becoming contains being and nothing in the sense that when something *becomes* it passes, as it were, from nothingness to being. But these contents cannot be understood apart from their contributions to the overarching category: this is what it is to be negated (*aufgehoben*) within the new category.

In general this is how the *Logic* proceeds: seeking its most basic and universal determination, thought posits a category to be reflected upon, finds then that this collapses due to a contradiction generated, like that generated by the category being, and so then seeks a further category with which to make retrospective senses of those contradictory categories. However, in turn the new category will generate some further contradictory negation and again the demand will arise for a *further*

concept that can reconcile these opposed concepts by incorporating *them* as moments.

The method Hegel employs here, *determinate negation*, is often compared with Spinoza's principle that "all determination is negation", but while Hegel's is *related* to Spinoza's thought, it cannot be identified with or reduced to it. Spinoza's top-down determination starts with a single category (in his case, *divine substance*) that is then progressively divided by the application of concepts—the model being Plato's method of division in which a *genus* concept is divided into particular *species* by the presence or absence of some differentiating property. From Hegel's point of view, however, this cannot capture *individuals* as other than *parts* of that greater whole—a metaphysical picture in relation to Spinoza he refers to as *acosmism*. So Hegel will balance *this* type of determination by negation, with a different type of negation modeled on that which holds between *incompatible properties* of some object (for example, red and blue as incompatible colors) and that is reflected in the *term negation* of Aristotle's logic. This allows Hegel to go beyond the determination of something as *particular* (suggesting the part-whole relation) to a more robust sense of *singularity* [*Einzelheit*]—the sense of the pure *thisness* seen initially in the *Phenomenology*'s Sense-certainty chapter, the truth of which was then shown to be Aristotle's idea of an individual thing's substantial form in the Perception chapter. It is in terms of this category that we can think, along with Aristotle, of a thing having an underlying substrate within which properties inhere and which, unlike the properties themselves, cannot be thought in *general* terms, but only in terms of the category of *singularity*. And yet this will encounter a problem for the *determinacy* of this underlying substrate—*it* will have to find determining contrasts that allow *it* to be determinately conceived. (In Book 2 of the *Logic* we will learn that the category of singularity will rely on particularity just as particularity has been shown to rely on singularlity.

Singular substrates or "essences" can only be known in relation to the general properties that constitute their appearances.)

Attempting to unravel the intricacies of the patterns of dependence between such categories will be task of this mammoth work, but here a general point might be made. If Hegel's thought here is considered to be, like Spinoza's, *holistic*, it is only so at a higher level of abstraction, such that these determinations of *singularity*, *particularity* and *universality* cannot themselves be understood in isolation from each other but only via their complex interactions. Hegel only *explicitly* explores the details of the interactions of these determinations of conceptuality in his discussion of judgments and syllogisms in Book 3, The Doctrine of Concept, suggesting that concerns of logic as traditionally conceived are not as irrelevant to the *Science of Logic* as often thought. However, the general point separating his approach from that of Spinoza clearly emerges earlier on. Determinate negation is not Spinoza's principle as Spinoza's presupposes a whole that *precedes* its parts, and that all *negations* are negations of something that is primitively *positive*. In contrast, Hegel's negations will be negations of determinations that are *already* to be conceived as *themselves* negations.

The other basic methodological principle of the *Logic* will be that this categorical infrastructure of thought is able to be unpacked using *only* the resources available to thought itself: the capacity of thought to make its contents determinate (in a way somewhat like what Leibniz had thought of as making clear but confused ideas clear and distinct), and its capacity to be consistent and avoid contradiction. Again, for some readers, this makes Hegel's logic akin to Kant's transcendental logic that, rather than treating the pure form of thought abstracted from *all* content, treats thought as already possessing a certain type of self-generated content, (in Kant's terminology, "transcendental content") that is presupposed by the subsequent acquisition of all *empirical* content. But if Hegel's is akin to Kant's transcendental logic, it has clear differences to it as well. For Kant,

transcendental logic was the logic governing the thought of finite thinkers like ourselves, whose cognition was constrained by the necessity of applying general discursive concepts to the singular contents given in sensory intuitions, and he contrasted this with the thought of a type of thinker not so constrained—God—a thinker whose thought could directly grasp the world in a type of intellectual intuition. While opinions divide as to how Hegel's approach to logic relates to that of Kant, it is important to grasp that for Hegel logic is not simply a science of the form of our thoughts. It is also a science of actual *content* as well, and as such has an ontological dimension.

The thought determinations of Book 1 lead eventually into those of Book 2, The Doctrine of Essence. Naturally the logical structures and processes implicit in essence-thinking are more developed than those of being-thinking. Crucially, the contrasting pair essence and appearance of Essence-logic allow the thought of some underlying reality that manifests itself *through* a different overlying appearance, in the way that the *forces* posited by the operations of *the Understanding* (explored in the *Phenomenology*'s Chapter 3) are grasped through the appearances they explain. In contrast, the categories of Being-logic seem to govern thought processes that are restricted to qualitative phenomena and their co-ordinations. But distinction between *essence* and *appearance* must itself instantiate the relation of determinate negation, and the metaphysical tendency to think of reality as made up of some underlying substrates in contrast to the superficial appearances will itself come to grief with the discovery that the notion of an essence is only meaningful in *virtue* of the appearance that it is meant to explain away. (In terms of the ultimate *conceptual* categories of singularity, particularity and universality, this discovery would be equivalent to grasping the idea that the *singularity* of the underlying, non-perceivable substrate or substantial form is meaningful only in relation to something that can bear the *particular* qualities that constitutes its worldly appearance.) For Hegel it is the

complex modern, but pre-Kantian, versions of substance metaphysics, like those of Spinoza and Leibniz, that bring out in the most developed way the inherently contradictory nature of this form of thought.

Book 3, The Doctrine of Concept, effects a shift from the Objective Logic of Books 1 and 2, to Subjective Logic, and metaphysically coincides with a shift to the modern subject-based category theory of Kant. Just as Kantian philosophy is founded on a conception of objectivity secured by conceptual coherence, Concept-logic commences with the concept of *concept* itself, with its moments of singularity, particularity and universality. While in the two books of objective logic, the movement had been between particular concepts, being, nothing, becoming etc., in the subjective logic, the conceptual relations are grasped at a meta-level, such that the concept *concept* treated in Chapter 1 of section 1 (Subjectivity) passes over into that of *judgment* in Chapter 2. It is important to grasp the basic contours of Hegel's treatment of judgment as it informs his subsequent treatment of inference.

Reprising an etymological point made by Hölderlin, Hegel notes that a judgment (*Urteil*) involves a separation (*Teilung*) of parts: in basic terms a predicate is said of some subject giving the judgment the grammatical form "*S is P*", but in *saying* "*S is P*", the judging subject affirms the *unity* existing between the parts. *S* and *P* are thus meant (1) to be diverse, but (2) to form a unity—a situation we are now familiar with in terms of the *Aufhebung* of parts in a whole. Hegel takes this as signaling two ways of thinking of the relation of subject and predicate in the judgment. One can take subject and predicate terms as self-subsistent entities that are *joined* in the judgment, or one can take the judgment itself as the primary unit that *splits* into subject and predicate terms. This in fact coincides with the two different ways in which logical relations have been conceived in the history of philosophy: the former represents the *term-logical* approach characteristic of Aristotle, while the latter represents the *propositional*

approach characteristic of the Stoics and much recent philosophy. From the former point of view one thinks of the subject term as designating a substance, typically grasped as an instance of a kind, in which properties, designated by predicate terms, *inhere*. From the latter point of view, one thinks of predicate terms as abstract universals that *subsume* or are satisfied by entities to which the subject terms refer, an approach which conceives of the propositional content, in Stoic terminology—the *lecton*, the *what-is-said*—as having a primacy over the parts. Using a distinction from the Medievals, we can describe the first type of judgments as *de re* (about things) and the second as *de dicto* (about sayings). These alternative *joining* and *splitting* approaches can in turn be applied to the relationship of *judgments* within *inferences* or *syllogisms*. While it is more common for inferences to be thought of as *composed* of judgments which have their own *truth values*, the judgments themselves can be thought of as gaining their meaning via the role they play in inferences, parallel to the way that the *parts* of the judgment can be thought of as resulting from the judgment's *splitting*. Within recent semantic theory, Robert Brandom has argued for such an *inferentialist* analysis and has suggested this way of understanding Hegel's logic (Brandom 2014), a view that fits with Hegel's idea that the *syllogism* is the "truth of the judgment" (SL: 593). Thought of in terms of the framework of Kant's transcendental logic, Hegel's position would be akin to allowing inferences—syllogisms—a role in the determination of the transcendental content of judgments, a role that is not allowed in Kant.

As we have said, Hegel's logic is meant somehow to generate a content—to *produce* a type of ontology—and this comes into explicit focus with Hegel's puzzling claim in Book 3 concerning a syllogism that has become "concrete" and "full of content" that thereby has *necessary* existence (SL: 616–7). In contrast with Kant, Hegel seems to go beyond a transcendental deduction of the *formal* conditions of experience and thought and to a deduction of their *material* conditions. Traditionalists will here point to

Hegel's allusions to the ontological proof (SL: 625) of medieval theology in which the *existence* of God is seen as necessitated by his concept—an argument undermined by Kant's criticism of the treatment of *existence* as a predicate. In Hegel's version, it is said, the objective existence that God achieves in the world is seen as necessitated by his essential self-consciousness. Non-traditional readings, in contrast, would have to interpret this aspect of Hegel's logic very differently. Brandom's inferentialist interpretation of Hegel, when joined to ideas taken from Hegel's treatment of self-consciousness in the *Phenomenology*, suggests a way forward here.

The first thing to be emphasized here is that we shouldn't think of judgments and their contents as something like *mental* contents—subjective or psychological states of a thinker's mind. Such a psychologistic attitude was opposed by Hegel just as it was opposed by a figure as central to modern logic as Gottlob Frege. For Frege, thoughts are not mental, rather they are *abstract* entities like numbers, so the problem facing us is not how to go from mental contents to the concrete world, it is how to go from *abstract* to concrete ones. But here we must keep in mind Hegel's two-fold way of thinking about judgments, *de dicto* and *de re*, and while it is usual to think of the contents of *de dicto* judgments as abstract (here to think of the content as *propositional* is usual), some have thought of the contents of *de re* judgments as including *the thing* itself (the "*re*") that the judgment is about. (In fact Bertrand Russell had, at points in his career, entertained such an idea of *propositional content* itself.) Thus when Hegel characterizes some judgment structures (typically perception based judgments) as *judgments of existence* one might take the perceived thing *itself* as straightforwardly part of the content of the judgment. It is a concrete object, but not grasped as a concrete *simple*, but grasped in relation to what is judged *of it* in the predicate. And to the extent that *judgments* can be considered components of syllogisms, we might

appreciate how syllogisms might have become contentful in a process that has culminated in the *concrete* syllogism of necessity.

If the concrete object of a *de re* judgment is effectively what had been under consideration in Chapter 2, Perception, in the *Phenomenology* (the thing with properties), we now might envisage where Hegel's thought is headed in these sections of the subjective logic. In the *Phenomenology* it turned out that the capacity for a subject to entertain objects of consciousness such as perceptual ones was that such a subject was capable of *self-consciousness*. It then turned out that to be capable of self-consciousness the subject had to exist in a world with other embodied subjects whose intentions it could recognize. It is here that we might pick up Robert Brandom's suggestion, following Sellars, that we should think of the existence of inferential processes or processes of reasoning as presupposing participation within social communicative interactions in which the making of an assertion is considered as a move in a language-game of the "giving and asking for reasons". In short, we may think of Hegel's syllogism of necessity, which constitutes the ground or "truth" of the earlier *formal* conception of syllogisms, as a type of inter-subjective practice embodying thought—a type of *syllogising* practice that is by necessity inter-subjective and recognitive. Formally considered we might think of this syllogism as the logical schematization of the most *developed* form of recognition in which thinkers acknowledge others *as* free thinkers.

I have suggested that in the syllogism of necessity with which Hegel's treatment of inference terminates we get a glimpse of a type of contentful and dynamic rational process unfolding in the midst of the recognitive and communicative interactions between finite living and intentional beings. What we see here is a reprise of the conception of logos as an objective process running through the world as had been conceived by the ancient Stoics and neo-Platonists. But it is now embedded not simply in the world as such—in *nature*—but in objectivized *spirit*, in human communities of

thinkers. We are now returned to the domain of *objectivity* that had characterized Books 1 and 2 of the *Science of Logic*, but we might expect such a return *from* subjectivity to have effected a change in *objectivity* as earlier understood.

To cross straight into a consideration of the objectivity of the human world of action and thought—spirit—would be to break the developmental pattern of the logic because thought about such a complex form of objective existence will presuppose thought about simpler forms. And so the starting point for the consideration of objectivity will again be that of the simple object as something immediately grasped by thought. But this object can now be developed with that elaborate conceptual apparatus that has emerged in the preceding section. Progression here will be from a naïve and immediate concept of an object as simple self-sufficient thing, a thing with its identity centered on itself, through the more complex idea of an object as grasped from within the interstices of physical and chemical thought, to the models of teleological and living systems. The *Logic* then transitions into a consideration of the "*adequate concept*, the objectively *true*, or the *true as such*" (SL: 670). This adequate concept is *the Idea*, which, after tracking through considerations of the living individual and theoretical and practical cognition, emerges as *the Absolute Idea*.

3.2 Hegel's *Encyclopaedic* System and its Expansions

As we have mentioned, Hegel's *Encyclopaedia of Philosophical Sciences* was written as a teaching manual, various parts of which were later expanded upon in lecture courses devoted to specific parts of the system. The first part of the *Encyclopaedia* is essentially a condensed version of his earlier *Science of Logic*, considered above. We will pass over a consideration of this work to the next component of the *Encyclopaedia*, Hegel's *Philosophy of Nature*.

3.2.1. Philosophy of Nature

Hegel's *Philosophy of Nature* (first published as such in 1842, and based on §§245–376 from the 1830 *Encyclopaedia*, supplemented by material and student transcripts from Hegel's Berlin lectures) has often been damned by the contention that Hegel had simply *dismissed* the activity of the natural sciences, especially Newtonian science, as based upon the inadequacies of the Understanding, and in their place had tried to somehow deduce the natural world from philosophical first principles. Recently, however, defenses of Hegel's philosophy of science have started to emerge, especially from the side of Hegel's reformed metaphysical interpreters. Thus, it has been argued by Westphal (2008) for example, that Hegel's philosophy of nature actually represents a sophisticated attempt to think through epistemological assumptions that are presupposed by the development of Newton's theory. Defending Hegel's philosophy of science from a similar point of view, James Kreines (2008) has argued for the relevance of Hegel's logical categories for the biological sciences of his times. We won't here attempt to present such arguments, but before any such reassessment of Hegel's work here *could* be undertaken, the fundamental criticism raised above of a project that attempts to base a philosophy of nature on his logic rather than the empirical sciences must be addressed. Was not Hegel simply trying to pre-empt the work of empirical scientists by somehow attempting to anticipate the very *contents* of their discoveries from logical considerations alone?

This objection is often summed up under the slogan of "deducing Krug's pen", in that in 1801 the philosopher W.T. Krug had accused *Schelling*'s idealist philosophy of nature of aiming to deduce the nature of all contingent phenomena, even that of the pen with which he, Krug, was writing his critique. Hegel responded to Krug's accusation in the following year, claiming that Krug had made the common mistake of conflating the

understanding with reason, and treating the Absolute as something on the same level as finite things.

Hegel was, at this time, closely aligned with Schelling's views, and would separate his own views *from* Schelling's in subsequent years leading up to the writing of *Phenomenology of Spirit*. Nevertheless, Hegel clearly thought that his point held regardless of the relation of his own views to Schelling's as he was to make similar points against Krug in a remark added to the *Philosophy of Nature* from the Heidelberg and Berlin periods. While logic must not be restricted to the "form" of an externally given "matter", nevertheless,

> it is the height of pointlessness to demand of the concept that it should explain … construe or deduce these contingent products of nature. (PN: §250, remark. Krug is mentioned explicitly in a footnote at this point.)

The point is expanded upon further when it is said that it is

> an error on the part of the philosophy of nature to attempt to face up to all phenomena; this is done in the finite sciences, where everything has to be reduced to general conceptions (hypotheses). In these sciences the empirical element is the sole confirmation of the hypothesis, so that everything has to be explained. (PN: §270, addition)

In keeping with the more general idea that that philosophy attempts to discern or recognize *concepts* in *representations* (*Vorstellungen*) or empirical appearances, philosophy of nature investigates the conceptual structures that are manifest in the products of the scientific work that is done on the basis of those appearances.

Traces of conceptual determination will certainly survive in the most particularized product, although they will not exhaust its nature. (PN: §250 remark)

Clearly, philosophy of nature is not in competition with the empirical natural sciences; it takes as its subject matter the results of those sciences in order to discover within them the particular ways in which the necessary categorial structures deduced in the logic are expressed.

In terms of topics treated, the *Philosophy of Nature* largely coincides with those treated in the third book of the *Science of Logic* when the logical processes and relations in question have returned to objectivity after the excursion into the *subjectivity* of formal logic at the outset of Book 3. In Mechanism Hegel had reconstructed a movement in thought from a primitive cosmology in which all objects are conceived in relation to a central object (the sun) that exemplifies objecthood *per se*, to a system of objects within which any such self-sufficient center has been eliminated. In this Newtonian world, that which gives order to the whole now has the ideality of law, but this is *itself* thought of as external to the system of objects.

After an Introduction, Section One of the *Philosophy of Nature*, Mechanics, expands on this progression through considerations of space and time, matter considered as the diversity of individual bodies distributed in space and time, and finally the idea of universal gravitation as the determinate concept of such corporeal matter realized as *idea* (PN: §270). In the Newtonian laws of mechanics, however, the unity of matter is still only *formal*, and in Section Two, Physics, the determinateness of form is now considered as *immanent within* such corporeal matter.

Matter has individuality to the extent that it is determined within itself by having being-for-self developed within it. It is through

this determination that matter breaks away from gravity and manifests itself as implicitly self-determining. (PN: §273)

While Mechanics clearly reflects the more space-filling conception of matter dominant in British thought, Physics is consistent with the more dynamic continental European conception of matter originating in Leibniz with his idea of living forces. Within this framework, Hegel attempts to organize a vast array of areas of contemporary physical investigation including meteorology, theories of sound and heat, light and electricity up to and including *chemical* processes which stand on the threshold of Organic Physics, dealt with in Section Three. The study of organics represents a return to the consideration of the individual *body* with which Mechanics had started, but now considered as "infinite process in which the individuality determines itself as the *particularity or finitude* which it also negates, and returns into itself by re-establishing itself at the end of the process as the beginning". The body is now "an *impregnated* and negative unity, which by relating itself to itself, has become essentially *self-centred and subjective*" (PN: §337). From such a conception, the first *body* to be considered is that of *the earth itself*, along with its history. Chapter Two moves to a consideration of the plant and Chapter Three, the animal organism.

From the point of view of the actual content of scientific theories and approaches that Hegel summarizes and locates within his system, his *Philosophy of Nature* is clearly a product of his time. Nevertheless, many of the underlying philosophical issues dealt with are still now far from settled. Thus, while Newtonian physics clearly became established in ways that made Leibniz's dynamic physics seem obsolete as *empirical* theory, debate still goes on as to whether conceptions of space-time in post-Newtonian physics is to be conceived in Newtonian or Leibnizian ways.

3.2.2 Philosophy of Subjective and Objective Spirit

In the *Encyclopaedia*, *Philosophy of Nature* is followed by *Philosophy of Spirit* (*Geist*). Hegel's usual triadic pattern when applied here results in sections devoted to the philosophies of subjective spirit, objective spirit, and absolute spirit. Philosophy of *subjective* spirit constitutes what is closest in Hegel's philosophy to a philosophy of mind in the contemporary sense, while the philosophy of *objective* spirit concerns those objective patterns of social interaction and the cultural institutions within which spirit is objectified in patterns of human life we have seen at work in *Phenomenology of Spirit*. Within subjective spirit, we may anticipate that the first division, Anthropology, will follow on from topics with which *Philosophy of Nature* ends—the animal organism—and so it does. Thus here Hegel is concerned with what he terms "*Seele*", "soul"—which seems to translate more the ancient Greek term, "psyche"—and hence the *mind-body* relation:

> If soul and body are absolutely opposed to one another as is maintained by the abstractive intellectual consciousness,

Hegel comments,

> then there is no possibility of any community between them. The community was, however, recognized by ancient metaphysics as an undeniable fact. (PN: § 389 add)

The *Seele* of Anthropology should therefore not be confused with the modern subjective conception of mind, as exemplified by Descartes and other early modern philosophers. Aristotle had conceived of the soul as the *form* of the body, not as a substance separate from that of the body, and had attributed lesser souls to animals and even plants. (Again, Aristotle's notion of *substantial form* comes into view.) Concomitantly, in this section Hegel describes spirit as *sunk* in nature, and treats consciousness as largely

limited to what now might be described as *sentient* or *phenomenal* consciousness alone—the feeling soul. *Consciousness* in the sense of the modern subject–object opposition only makes its appearance in the following second section, Phenomenology of Spirit, which, reprising key moments from the earlier book of that name, raises a problem for how we are to understand the relation of phenomenology and systematic philosophy: is it a *path* to it or *part* of it? Given that the recognitive approach to self-consciousness *presupposes* that potential self-consciousnesses are *in fact* embodied and located in the world, we would expect the mind as treated in Psychology to be no less embodied as the way in which it is conceived in Anthropology. What in fact distinguishes the mind of Psychology from that of Anthropology is its rational capacities, considered in terms that would now be described as normative rather than simply naturalistic, and this for Hegel clearly signals a difference in the way in which an actual psychological subject *relates to* his or her own body. The type of *abstractive* thinking found in Psychology does not, of course, as in mythical images of metempsychosis—a favorite trope of Platonists—involve the mind leaving the body. This would count for Hegel as a piece of mythical picture thinking—a *Vorstellung*. Rather, it involves a certain capacity of the psychological subject to suspend unreflected-upon endorsement of the *claims* made on behalf of his or her body, for example, to subject the evidence given by the senses to rational scrutiny.

Given the dialectical mode in which Hegel's texts progress, as seen already in both *Phenomenology of Spirit* and *Science of Logic*, we will expect the capacities examined in Psychology to ultimately depend upon those that come under consideration in the context of *objective* spirit. In this sense, we are witnessing within another mode, the type of progression seen in the movement in Phenomenology from shapes of consciousness to shapes of spirit. The internal *Phenomenology of Spirit* seems to play an important role in setting up this transition from Psychology to *Objective*

Spirit (Williams 2007), but it might also be seen as crucial in relating the more cognitive dimensions of Psychology back to the theme of embodiment prominent in Anthropology (Nuzzo 2013a).Thus any *naturalistic analysis* is ultimately surpassed by a social and historical one, which *itself* cannot be understood as *anti*-naturalistic.

The philosophy of *subjective* spirit passes over into that of *objective spirit*, which concerns the objective patterns of social interaction and the cultural institutions within which spirit is objectified. The book entitled *Elements of the Philosophy of Right*, published in 1821 as a textbook to accompany Hegel's lectures at the University of Berlin, essentially corresponds to a more developed version of the philosophy of objective spirit and will be considered here.

Elements of the Philosophy of Right

The *Philosophy of Right* (as it is more commonly called) can be read as a political philosophy that stands independently of the system (Tunick 1992), despite the fact that Hegel intended it to be read against the background of the developing conceptual determinations of the *Logic*. The text proper starts from the conception of a singular willing subject (grasped from the point of view of its individual self-consciousness) as the bearer of *abstract* right. While this conception of the individual willing subject possessing some kind of fundamental rights was in fact the starting point of many modern political philosophies (such as that of Locke, for example) the fact that Hegel commences here does not testify to any ontological assumption that the consciously willing and right-bearing individual is the basic atom from which all society can be understood as constructed—an idea at the heart of standard social contract theories. Rather, this is simply the most immediate starting point of Hegel's presentation and corresponds to analogous starting places of the

Phenomenology and the *Logic*. Just as the categories of the *Logic* develop in a way meant to demonstrate that what had at the start been conceived as simple is in fact only made determinate in virtue of its being a functional part of some larger structure or process, here too it is meant to be shown that any simple willing and right-bearing subject only gains its determinacy in virtue of a place it finds for itself in a larger social, and ultimately historical, structure or process. Thus, even a contractual exchange (the minimal social interaction for contract theorists) is not to be thought simply as an occurrence consequent upon the existence of two beings with natural animal wants and some natural calculative rationality, as in Hobbes, say; rather, the system of interaction within which individual exchanges take place (the economy) will be treated holistically as a culturally-shaped form of social life within which the actual wants of individuals as well as their reasoning powers are given determinate forms. Hegel is well aware of the distinctive modernity of this form of social-life.

Here too it becomes apparent that Hegel, taking up themes from the Phenomenology, follows Fichte in treating property in terms of a *recognitive* analysis of the nature of such a right. A contractual exchange of commodities between two individuals itself involves an implicit act of recognition in as much as each, in giving something to the other in exchange for what they want, is thereby recognizing that other *as* a proprietor of that thing, or, more properly, of the inalienable *value* attaching to it (PR: §§72–81). By contrast, such proprietorship would be *denied* rather than recognised in fraud or theft—forms of wrong (*Unrecht*) in which right is negated rather than acknowledged or posited (§§82–93). Thus what differentiates property from mere *possession* is that it is grounded in a relation of reciprocal recognition between two willing subjects (§71 and remark). Moreover, it is in the exchange relation that we can see what it means for Hegel for individual subjects to share a *common will*—an idea which will have important implications with respect to the difference of Hegel's conception of the state from that of Rousseau. Such

an interactive constitution of the common will means that for Hegel that the identity among wills is achieved *because of* not *in spite of* co-existing differences between the particular wills of the subjects involved: while contracting individuals both will *the same* exchange, at a more concrete level, they do so with different ends in mind. Each wants something different *from* the exchange. Without this difference, the type of absorption of individual subjects into collective *substance* of the type of Hegel worries about in relation to Spinoza would occur (§258 remark).

Hegel passes from the abstractly individualistic frame of Abstract Right to the social determinacies of *Sittlichkeit* or Ethical Life (PR: §142) via considerations first of *wrong*, the negation of right, (§§82–96) and the punishment that such wrong entails the negation of wrong, and hence the "negation of the negation" of the original right (§§97–104), and then of *morality*, conceived more or less as an internalisation of the external legal relations presupposed by punishment. Consideration of Hegel's version of the retributivist approach to punishment affords a good example of his use of the logic of negation. In punishing the criminal the state makes it clear to its members that it is the acknowledgment of right *per se* that is essential to developed social life: the significance of acknowledging another's right in the contractual exchange cannot be, as it at first might have appeared to the participants, simply that of an instrumental means by which each gets what he or she wants from the other.

Hegel's treatment of punishment also brings out the continuity of his way of conceiving of the structure and dynamics of the social world with that of Kant, as Kant too, in his *Metaphysics of Morals* had employed the idea of the state's punitive action as a *negating* of the original criminal act. Kant's idea, conceived on the model of the physical principle of action and reaction, was structured by the category of community or reciprocal interaction, and was conceived as involving what he called *real opposition*. Such an idea of opposed dynamic forces seems to form

something of a model for Hegel's idea of contradiction and the starting point for his conception of reciprocal recognition. Nevertheless, clearly Hegel articulates the structures of recognition in more complex ways than those derivable from Kant's category of community.

First of all, in Hegel's analysis of *Sittlichkeit* the type of sociality found in the market-based civil society is to be understood as dependent upon and in contrastive opposition with the more immediate form found in the institution of the family: a form of sociality mediated by a quasi-natural inter-subjective recognition rooted in sentiment and feeling—love (PR: §§ 158–60). (This dependence shows how *anthropological* determinations do not simply disappear with the development of more *psychological* ones— they are *preserved* as well as *negated* as in the pattern of what is *aufgehoben*. It also shows the mutual dependence of the determinations of the *singularity* of the atomistic subjects of civil society and their *particularity* as members (parts) of holistically conceived families.) Here Hegel seems to have extended Fichte's legally characterized notion of recognition into the types of human inter-subjectivity earlier broached by Hölderlin and the romantics. In the family the particularity of each individual tends to be absorbed into the social unit (one is a *part of* one's family), giving this manifestation of *Sittlichkeit* a one-sidedness that is the inverse of that found in market relations in which participants grasp themselves in the first instance as *singular* [*einzeln*] self-identical individuals who *then* enter into relationships that are external to them.

These two opposite but interlocking principles of social existence provide the basic structures in terms of which the component parts of the modern state are articulated and understood. As both contribute particular characteristics to the subjects involved in them, part of the problem for the rational state will be to ensure that each of these two principles mediates the other, each thereby mitigating the *one-sidedness* of the other. Thus, individuals who encounter each other in the external relations of the

market place and who have their subjectivity shaped by such relations also belong to families where they are subject to opposed influences. Moreover, even within the ensemble of production and exchange mechanisms of civil society individuals will belong to particular *estates* (the agricultural estate, that of trade and industry, and the *universal* estate of civil servants (PR: §§199–208)), whose internal forms of sociality will show family-like features.

Although the actual details of Hegel's mapping of the categorical structures of the *Logic* onto the *Philosophy of Right* are far from clear, the general motivation is apparent. Hegel's logical categories can be read as an attempt to provide a schematic account of the material (rather than formal) conditions required for developed self-consciousness for whom rationality and freedom are maximized. Thus we might regard the various syllogisms of Hegel's *Subjective Logic* as attempts to chart the skeletal structures of those different types of recognitive inter-subjectivity necessary to sustain various aspects of rational cognitive and conative functioning (self-consciousness). From this perspective, we might see his logical schematisation of the modern rational state as a way of displaying just those sorts of mediating institutions that a state must provide if it is to answer Rousseau's question of the form of association needed for the formation and expression of the *general will*.

Concretely, for Hegel it is representation of the estates within the legislative bodies that is to achieve this (PR: §§301–14). As the estates of civil society group their members according to their common interests, and as the deputies elected from the estates to the legislative bodies give voice to those interests within the deliberative processes of legislation, the outcome of this process might give expression to the general interest. But Hegel's republicanism is here balanced by his invocation of the familial principle: such representative bodies can only provide the *content* of the legislation to a *constitutional monarch* who must add to it the form of the

royal decree—an individual "I will …" (§§275–81). To declare that for Hegel the monarch plays only a symbolic role here is to miss the fundamentally idealist complexion of his political philosophy. The expression of the general will in legislation cannot be thought of as an outcome of some quasi-mechanical process: it must be *willed*. If legislation is to express the general will, citizens must recognize it as expressing *their* wills; and this means, recognising it *as* willed. The monarch's explicit "I will" is thus needed to close this recognitive circle, lest legislation look like a mechanical compromise resulting from a clash of contingent interests, and so as actively *willed* by nobody. Thus while Hegel is critical of standard social contract theories, his own conception of the state is still clearly a complicated transformation of those of Rousseau and Kant.

Perhaps one of the most influential parts of Hegel's *Philosophy of Right* concerns his analysis of the contradictions of the unfettered capitalist economy. On the one hand, Hegel agreed with Adam Smith that the interlinking of productive activities allowed by the modern market meant that "subjective selfishness" turned into a "contribution towards the satisfaction of the needs of everyone else" (PR: §199). But this did not mean that he accepted Smith's idea that this "general plenty" produced thereby *diffused* (or "trickled down") though the rest of society. From within the type of consciousness generated within civil society, in which individuals are grasped as *bearers of rights* abstracted from the particular concrete relationships to which they belong, Smithean optimism may seem justified. But this simply attests to the one-sidedness of this type of abstract thought, and the need for it to be mediated by the type of consciousness based in the family in which individuals are grasped in terms of the way they *belong to* the social body. In fact, the unfettered operation of the market *produces* a class caught in a spiral of poverty. Starting from this analysis, Marx later used it as evidence of the need to abolish the individual proprietorial rights at the heart of Hegel's civil

society and socialise the means of production. Hegel, however, did not draw this conclusion. His conception of the exchange contract as a form of recognition that played an essential role within the state's capacity to provide the conditions for the existence of rational and free-willing subjects would certainly prevent such a move. Rather, the economy was to be contained within an over-arching institutional framework of the state, and its social effects offset by welfarist intervention. Some of Hegel's most telling criticisms of the unmediated effects of modern civil society concern those on the psychological lives of individuals. Recently, an approach to social reality with Hegelian provenance that uses the notion of recognition to articulate such pathologies has been developed by Axel Honneth (2010), testifying to the continuing relevance of Hegel's analyses.

Philosophy of History

The final 20 paragraphs of the *Philosophy of Right* (and the final 5 paragraphs of objective spirit section of the *Encyclopaedia*) are devoted to *world history* (*die Weltgeschichte*), and they also coincide with the point of transition from objective to *absolute* spirit. We have already seen the relevance of historical issues for Hegel in the context of the *Phenomenology of Spirit*, such that a series of different forms of objective spirit can be grasped in terms of the degree to which they enable the development of a universalizable self-consciousness capable of rationality and freedom. Hegel was to enlarge on these ideas in a lecture series given five times during his Berlin period, and it was *via* the text assembled on the basis of these lectures by his son Karl, that many readers would be introduced to Hegel's ideas after his death.

World history is made up of the histories of *particular* peoples within which spirit assumes some "particular principle on the lines of which it

must run through a development of its consciousness and its actuality" (PM: §548). Just the same dialectic that we have first seen operative among shapes of consciousness in the *Phenomenology* and among categories or thought-determinations in the *Logic* can be observed here. An historical community *acts* on the principle that informs its social life, the experience and memory of this action and the consequences it brings —a memory encoded in the stories that circulate in the community— results in this principle becoming available for the self-consciousness of the community, thus breaking the *immediacy* of its operation. This loss of immediacy brings about the decline of that community but gives rise to the principle of a new community:

> in rendering itself objective and making this its being an object of thought, [spirit] on the one hand destroys the determinate form of its being, and on the other hand gains a comprehension of the universal element which it involves, and thereby gives a new form to its inherent principle … [which] has risen into another, and in fact a higher principle. (PWH: 81)

This dialectic linking concrete communities into a developmental narrative which shows

> the path of liberation for the spiritual substance, the deed by which the absolute final aim of the world is realized in it, and the merely implicit mind achieves consciousness and self-consciousness. (PM: §549)

It is a dialectic, however, which only passes through *some* communities. Hegel's is a clearly *Eurocentric* account of history.

It is thus that "the analysis of the successive grades [of universal history] in their abstract form belongs to logic" (PWH: 56), but once more, it must be stressed that, as with philosophy of nature, philosophy of history is not

meant to somehow magically *deduce* actual empirical historical phenomena, like Krug's pen; rather, it takes the results of actual empirical history as its material and attempts to find exemplified within this material the sorts of categorial progressions of the logic. Thus the activity of the philosophical historian presupposes that of "original" and "reflective" historians (PWH: 1–8). The actual world is full of contingencies from which empirical historians will have already abstracted in constructing their narratives, for example, when writing from particular *national* perspectives. To grasp history *philosophically*, however, will be to grasp it from *the perspective of world-history* itself, and this provides the transition to *absolute* spirit, as world history will understood in terms of the manifestation of what from a religious perspective is called "God", or from a philosophical perspective, "reason". Hegel clearly thinks that there is a way of cognitively relating to history in a way that goes beyond the standpoint of consciousness and the understanding—the standpoint of what we now think of as informing scientific history. From the perspective of consciousness history is something that stands over against me *qua* something known, but from the standpoint of *self*-consciousness I grasp this history as the history of that which contributes to *me*, qua rational and free being.

3.2.3 Philosophy of Absolute Spirit

The subject matter of the final 25 paragraphs of the *Encyclopaedia Philosophy of Spirit*, Absolute Spirit, came to be expanded massively into the contents of three different lectures series on philosophy of art, religion, and history of philosophy given multiple times during Hegel's decade in Berlin. Assembled and published in the years immediately following his death, these were the works through which Hegel was to become known as perhaps the most significant synoptic theorist of these cultural phenomena. Rather than to attempt to capture the richness of his thought here in a few paragraphs, which would be bound to be futile, I will simply

try to allude to how this material is meant to draw upon the conceptual resources noted so far.

Art

(See also, Hegel's aesthetics)

Hegel was writing in a time of intense development of ideas about the arts. Kant had treated aesthetic experience largely in relation to the experience of the beauty of nature, but for Hegel aesthetics becomes primarily about *art*. The reason for this is simple: art is an objective medium in which a community collectively reflects upon itself, and the art of historical peoples is to be understood in as the attempt to bring before the consciousnesses of its members the totality of what is. It is *as* art that "consciousness of the Absolute first takes shape" (PM: §556). The peculiarity of art lie in the *sensuousness* of the medium in which its content is objectified.

In the 1790s, Friedrich Schiller and Friedrich Schlegel had given aesthetics an historical dimension, distinguishing the forms of ancient and modern art in terms of the contrasts naïve–sentimental and classical–romantic respectively. Hegel adopts Schlegel's terminology to distinguish as *classical* the art that thrived in the Greek and Roman worlds from the *romantic* art of post-classical times. Again, the romantic or modern here will be characterized by the depth of a form of individual *subjective* consciousness that is largely missing in antiquity. But those in Greek antiquity, where *psychological* determinations were closer to *anthropological* ones, had lived with a comfortable felt unity between spirit and body and between the individual and society. A characteristic of the Greeks was their *Heimatlichkeit*—their collective feeling of being *at home* in the world as they were each at home in their bodies. Modern subjectivity is thereby purchased as the expense of a sense of *abstraction*

and *alienation* from the actual world and from the self—a consequence of the way the modern subject has become related to his or her body in a different way.

Hegel, influenced by the work of a former colleague, the Heidelberg philologist Friedrich Creuzer, adds to Schlegel's categorization of art forms by positing a further category of *symbolic* art, characterizing the material cultures of ancient Eastern civilizations such as Persia, India and Egypt. The symbolic art of pantheistic religions of the East used *natural elements* to symbolize the gods of their cultures: Zoroastrianism had taken light, for example, to symbolize the divine (Aes I: 325), and animal worship was found in the Egyptians (Aes I: 357). But such actual *things* had to be distinguished from what was meant to be symbolized by them, so violence had to be done to such natural forms in attempts to represent the absolute—such cultural products thus becoming "bizarre, grotesque, and tasteless" (Aes I: 77), This, however, undermined their initial function, and the Greeks were able to offer a dialectical solution to this contradiction. They gave expression to the Absolute or the Idea by taking as its material the specifically *human form*, but only on condition of its being rendered "exempt from all the deficiency of the purely sensuous and from the contingent finitude of the phenomenal world". But even as idealized in Greek sculpture, say, the represented Greek god is still an object of "naïve intuition and sensuous imagination" (Aes I: 77–8), and as such the classical gods contained the germ of their own decline as they could not evade

> the finitudes incidental to anthropomorphism [which] pervert the gods into the reverse of what constitutes the essence of the substantial and Divine (Aes I: 502–4)

A new form of art will be needed to resolve these contradictions, and this is provided by romantic art. But the *material* for this form will not come

from within art itself. While Greek art can be understood as simultaneously belonging to aesthetic and religious realms, romantic art results from a fission within the symbolic realm of what in the *Phenomenology* Hegel had treated as a single category, Art-Religion. The transition from classical art to romantic art represents both a liberation of art from religion and of religion from art and the sensuous. Thus Christianity, whose rituals centered around the myth of God becoming man in the person of Jesus, had avoided the type of reliance on the beautiful productions of art in the way that characterized Greek religions. The shift from classical to romantic art, then, represents a broader shift between a culture whose final authority was an aesthetic one and a culture in which this authority was handed over to religion, and thus represents a shift in the authoritativeness of different cognitive forms. This loss of ultimate authority is the meaning of Hegel's often misunderstood thesis of *the death of art*.

Religion

It is well known that after Hegel's death in 1831, his followers soon split into left, centre and right factions over the issue of religion. A dispute over an appropriately Hegelian philosophical attitude to religion had been sparked by the publication in 1835–6 of David Strauss's *The Life of Jesus Critically Examined*—the conservative *right* claiming that Hegelianism reflected Christian orthodoxy, the *left* seeing it as a humanistic doctrine concerning the historical emancipation of mankind. In fact the implications of Hegel's philosophy for religious belief had been contentious since his rise to prominence in the 1820s. While officially declaring that philosophy and religion had the same *content*—God—Hegel claimed that the *conceptual* form of philosophy dealt with this concept in a more developed way than that which was achievable in the imagistic representational form of religion. Many opponents were suspicious that

the concept of God was emptied of its proper meaning in the process of Hegel's philosophical translations and Hegel was suspected by some of pantheism or atheism. Ultimately, then, the source of the corrosive effects of Hegel's philosophy on religion indeed could appear to be the insistence that the content of religious belief, like everything else, be grounded on rational, in fact *logical*, considerations—the logical coherence of the system of philosophy itself—rather than on anything like revelation.

In the writings he had produced in the 1790s Hegel had shown a clear attraction to the type of folk art-religions of ancient Greece in contrast to Christianity, whose other-worldly doctrines did not reflect the kind of *Heimatlichkeit* he valued in the ancient world, and it is common to see Hegel's later embrace of the Christianity he described as "the consummate religion" as an expression of a cultural and political conservatism of his later years. This under-estimates the complexity of Hegel's evolving views on both philosophy and religion, however. The limitations of Greek *at-homeness* in the world had to do with the inability of Greek life and thought to sustain that dimension of human existence that is reflected in the category of *singularity* of the subject. The fate of Socrates had thus represented the ultimate incompatibility with the Greek form of life itself of the type of individual, reflective individual who could reflectively bring *any* belief into question and take a stand against convention. Similar incompatibilities could be seen reflected in Greek tragedies such as *Antigone*.

With the decline of the Greek world and the rise of the Roman one, movements such as Stoicism and Christianity would come to give expression to an individual point of view, but under the social conditions of Rome or the Middle Ages such a subjective point of view could only be an alienated one attracted to what, in contrast to Greek concreteness, would be seen as abstractions. Prior to the modern world there would be no *real* place in either everyday life or in philosophical culture for any

non-alienated versions of the reflective or subjective position that had first emerged with Socrates—no form of life in which this individual dimension of human subjectivity *could be* at home. But Christianity marked a type of advance over Stoicism in that its doctrines of the nature of a good life had a this worldly *exemplar*. Thus in describing the doctrine of the Stoic sage, Hegel seemed fond of quoting Cicero's dictum that *nobody can say who this sage is* (LHP II: 250–1, 256). It was the abstractness of the Stoic's conception of the good man that was answered by the new religious cult centred on the life of the historical Christ.

In this sense Christianity marked a definite advance over the more intuitively based religious cults to which Hegel had been attracted in his youth, but it would only be in the modern world that the content of the core ideas of Christianity could be given proper expression. Thus Hegel treats medieval Catholicism as still caught in the abstractions of a transcendent realm and as caught up in a type of literal reading of this religion's *Vorstellungen*—its pictures. These need to become conceptualized, and this happens under modern Protestantism, and this, for Hegel, requires a type of demythologization of the religious content handed down from the past. Christ must somehow come to stand as an example of the human kind in general, which is the ultimate bearer of the status of being the *son of God*. Once more, it is the purported *singularity* of the category *son of God* that must be brought back into relation to the *universality* of the human *genus*. It is the nature of this result that divided Hegel's followers into their right and left camps. The understanding of what Hegel means by the concept *religion* in turn becomes tied to understanding what he means by *philosophy*. Appropriately the *Philosophy of Religion* thus passes over into the final form of *Absolute Spirit*, Philosophy—a science that is the "unity of Art and Religion" (PM: §572). The mere six paragraphs devoted to this science in the *Encyclopaedia* and dealing almost exclusively with the relation of philosophy to religion were to be expanded into the massive posthumously

published three volumes on the (philosophical) history of philosophy based on various sources including student transcripts for his lecture series given in Berlin.

History of Philosophy

In Hegel's time, the idea that philosophy *had* a historical development had only recently come into focus. Both Fichte and Schelling had discussed the idea of a history of reason following Kant's allusion to such a notion in the closing pages of *Critique of Pure Reason*, and systematic approaches to the history of philosophy had emerged like that of W.G. Tennemann, who presupposed a type of Kantian framework. Clearly Hegel's history of philosophy would be a present-centered one, in which the philosophical narrative would reveal a development up to the point represented by his own philosophy as *its* culmination. It is thus predictably Eurocentric: philosophy "commences in the West" because the West is where "this freedom of self-consciousness first comes forth" (LHP I: 99). There is an important caveat to add here, however. Philosophy is often identified with the capacity for *abstract* thought, and this is not confined to Europe and its history. Rather, it is typical of *eastern* cultures like those of India and China. As we have seen in the context of *art*, Hegel identifies Greek culture with a type of at-homeness in the world—what we might think of as the *opposite* of a tendency to abstraction and its typical attraction to the *transcendent* or *other-worldly*.

Greek philosophy, and so *philosophy itself*, starts with Thales and Ionian natural philosophy. When Thales choses *water* as the "the principle and substance of all that is" (LHP I: 175) he has abstracted the *concept* water from the stuff immediately encountered in puddles and so on. Subsequent attempts to specify what it is that underlies all things show influences of eastern abstraction as in Pythagoras's numericism, which is static and

"destitute of process or dialectic" (LHP I: 212) but later thinkers such as Zeno and Heraclitus grasp that which is at the heart of things in more dynamic ways. This type of dialectical thought which grasps the *unity* of opposites, familiar from the *Phenomenology* and the *Logic*, comes to fruition in Plato's dialogue *Parmenides* (LHP I: 261). What we are witnessing here, of course, is meant to be a progression that in some sense mirrors the progression of categories in Hegel's own *Logic*, but this progression of objective *content* gets joined to another dynamic with the appearance of Socrates.

Socrates was more than a philosopher: he was a "world-historical person"—a "main turning-point [*Hauptwendepunkt*] of spirit on itself" exhibited itself in his philosophical thought (LHP I: 384). In short, Socrates had added a *subjective dimension* to the otherwise natural moral lives of Athenian citizens, in that he had challenged them to find the principles not of worldly things but of *their own actions*, and challenged them to find these within the resources of *their own individual consciousnesses*.

> In him we see pre-eminently the inwardness of consciousness that in an anthropological way existed in the first instance in him and became later on a usual thing. (LHP I: 391)

With this we see "moral substance [*Sittlichkeit*] turn into reflective morality [*Moralität*]" and "the reflection of consciousness into itself". "The spirit of the world here begins to turn, a turn that was later carried to its completion" (LHP I: 407). This completion would be only achieved in modernity because, as we have seen, Socrates's challenging of convention in terms of resources taken from individual consciousness itself was incompatible with the immediacy of Greek *Sittlichkeit*.

Plato and, especially Aristotle, represent the pinnacle of ancient philosophy, but this philosophy, no matter how great, represents *its time*, that is, the time of the Greek form of spirit, raised to the level of thought. Neither Plato nor Aristotle can break free in thought from the contradiction between the conception of autonomous subjectivity represented by Socrates and the essential collectivity of Greek culture. Classical Greek philosophy will succumb in the same way that the Greek polis succumbs to its own internal contradictions, and what will eventually replace it will be a type of philosophizing constrained within the doctrinal constraints of the new religion, Christianity. But Christianity, as we have seen, gives representation to a solution to the problem of subjectivity encountered in the form of Socrates.

Philosophy proper only thrives under conditions of at-homeness in the world and such conditions obtained in neither the Roman nor medieval world. Hegel then sees both periods of philosophy as effectively marking time, and it is only in the modern world that once more develops. What modern philosophy will reflect is the universalization of the type of subjectivity we have seen represented by Socrates in the Greek polis and Jesus in the Christian religious community. Strangely, Hegel nominates two very antithetic figures as marking the onset of modern philosophy, Francis Bacon and the German Christian mystic, Jacob Böhme (LHP III: 170–216). In the 1825–6 lectures, from there Hegel traces the path of modern philosophy through three phases: a first period of metaphysics comprising Descartes, Spinoza and Malebranche; a second treating Locke, Leibniz and others; and the *recent* philosophies of Kant, Fichte, Jacobi and Schelling. Of course the perspective from which this narrative has been written is the absent *final* stage within this sequence—that represented by Hegel himself. Hegel concludes the lectures with the claim that he has

> tried to exhibit their (this series of spiritual configurations) necessary procession out of one another, so that each philosophy

necessarily presupposes the one preceding it. Our standpoint is the cognition of spirit, the knowledge of the idea as spirit, as absolute spirit, which as absolute opposes itself to another spirit, to the finite spirit. To recognize that absolute spirit can be *for it* is this finite spirit's principle and vocation. (LHP 1825–6, III: 212)

Bibliography

German Works

Gesammelte Werke. Edited by the Rheinisch-Westfälischen Akademie der Wissenschaften. Hamburg: Felix Meiner Verlag, 1968–.

Vorlesungen: Ausgewählte Nachschriften und Manuskripte. Edited by Pierre Garniron and Walter Jaeschke. Hamburg: Felix Meiner Verlag, 1983–.

Werke in zwanzig Bänden. Edited by Eva Moldenhauer and Karl Markus Michel. Frankfurt am Main: Suhrkamp Verlag, 1971.

English Translations of Key Texts

[ETW], *Early Theological Writings*, translated by T.M. Knox, Chicago: Chicago University Press, 1948.

The Difference Between Fichte's and Schelling's System of Philosophy, translated by H.S. Harris and W. Cerf, Albany: State University of New York Press, 1977.

[Phen], *Phenomenology of Spirit*, translated by A.V. Miller, Oxford: Oxford University Press, 1977.

[SL], *The Science of Logic*, translated by George di Giovanni, New York: Cambridge University Press, 2010.

The Encyclopedia Logic: Part 1 of the Encyclopaedia of Philosophical Sciences, translated by T.F. Geraets, W.A. Suchting, and H.S. Harris, Indianapolis: Hackett, 1991.

Ludwig Wittgenstein

First published Fri Nov 8, 2002; substantive revision Wed May 2, 2018

Considered by some to be the greatest philosopher of the 20th century, Ludwig Wittgenstein played a central, if controversial, role in 20th-century analytic philosophy. He continues to influence current philosophical thought in topics as diverse as logic and language, perception and intention, ethics and religion, aesthetics and culture. Originally, there were two commonly recognized stages of Wittgenstein's thought—the early and the later—both of which were taken to be pivotal in their respective periods. In more recent scholarship, this division has been questioned: some interpreters have claimed a unity between all stages of his thought, while others talk of a more nuanced division, adding stages such as the middle Wittgenstein and the third Wittgenstein. Still, it is commonly acknowledged that the early Wittgenstein is epitomized in his *Tractatus Logico-Philosophicus*. By showing the application of modern logic to metaphysics, via language, he provided new insights into the relations between world, thought and language and thereby into the nature of philosophy. It is the later Wittgenstein, mostly recognized in the *Philosophical Investigations*, who took the more revolutionary step in critiquing all of traditional philosophy including its climax in his own early work. The nature of his new philosophy is heralded as anti-systematic through and through, yet still conducive to genuine philosophical understanding of traditional problems.

1. Biographical Sketch

Wittgenstein was born on April 26, 1889 in Vienna, Austria, to a wealthy industrial family, well-situated in intellectual and cultural Viennese circles. In 1908 he began his studies in aeronautical engineering at Manchester University where his interest in the philosophy of pure mathematics led him to Frege. Upon Frege's advice, in 1911 he went to Cambridge to study with Bertrand Russell. Russell wrote, upon meeting Wittgenstein: "An unknown German appeared … obstinate and perverse, but I think not stupid" (quoted by Monk 1990: 38f). Within one year, Russell was committed: "I shall certainly encourage him. Perhaps he will do great things … I love him and feel he will solve the problems I am too old to solve" (quoted by Monk 1990: 41). Russell's insight was accurate. Wittgenstein was idiosyncratic in his habits and way of life, yet profoundly acute in his philosophical sensitivity.

During his years in Cambridge, from 1911 to 1913, Wittgenstein conducted several conversations on philosophy and the foundations of logic with Russell, with whom he had an emotional and intense relationship, as well as with Moore and Keynes. He retreated to isolation in Norway, for months at a time, in order to ponder these philosophical problems and to work out their solutions. In 1913 he returned to Austria and in 1914, at the start of World War I (1914–1918), joined the Austrian army. He was taken captive in 1918 and spent the remaining months of the war at a prison camp. It was during the war that he wrote the notes and drafts of his first important work, *Tractatus Logico-Philosophicus*. After the war the book was published in German and translated into English.

In 1920 Wittgenstein, now divorced from philosophy (having, to his mind, solved all philosophical problems in the *Tractatus*), gave away his part of his family's fortune and pursued several 'professions' (gardener, teacher, architect, etc.) in and around Vienna. It was only in 1929 that he returned to Cambridge to resume his philosophical vocation, after having been exposed to discussions on the philosophy of mathematics and science with members of the Vienna Circle, whose conception of logical empiricism was indebted to his *Tractatus* account of logic as tautologous, and his philosophy as concerned with logical syntax. During these first years in Cambridge his conception of philosophy and its problems underwent dramatic changes that are recorded in several volumes of conversations, lecture notes, and letters (e.g., *Ludwig Wittgenstein and the Vienna Circle*, *The Blue and Brown Books*, *Philosophical Grammar*). Sometimes termed the 'middle Wittgenstein', this period heralds a rejection of dogmatic philosophy, including both traditional works and the *Tractatus* itself.

In the 1930s and 1940s Wittgenstein conducted seminars at Cambridge, developing most of the ideas that he intended to publish in his second book, *Philosophical Investigations*. These included the turn from formal logic to ordinary language, novel reflections on psychology and

mathematics, and a general skepticism concerning philosophy's pretensions. In 1945 he prepared the final manuscript of the *Philosophical Investigations*, but, at the last minute, withdrew it from publication (and only authorized its posthumous publication). For a few more years he continued his philosophical work, but this is marked by a rich development of, rather than a turn away from, his second phase. He traveled during this period to the United States and Ireland, and returned to Cambridge, where he was diagnosed with cancer. Legend has it that, at his death in 1951, his last words were "Tell them I've had a wonderful life" (Monk: 579).

2. The Early Wittgenstein

2.1 *Tractatus Logico-Philosophicus*

Tractatus Logico-Philosophicus was first published in German in 1921 and then translated—by C.K. Ogden, with F. P. Ramsey's help—and published in English in 1922. It was later re-translated by D. F. Pears and B. F. McGuinness. Coming out of Wittgenstein's *Notes on Logic* (1913), "Notes Dictated to G. E. Moore" (1914), his *Notebooks*, written in 1914–16, and further correspondence with Russell, Moore and Keynes, and showing Schopenhauerian and other cultural influences, it evolved as a continuation of and reaction to Russell and Frege's conceptions of logic and language. Russell supplied an introduction to the book claiming that it "certainly deserves … to be considered an important event in the philosophical world." It is fascinating to note that Wittgenstein thought little of Russell's introduction, claiming that it was riddled with misunderstandings. Later interpretations have attempted to unearth the surprising tensions between the introduction and the rest of the book (or between Russell's reading of Wittgenstein and Wittgenstein's own self-assessment)—usually harping on Russell's appropriation of Wittgenstein for his own agenda.

The *Tractatus*'s structure purports to be representative of its internal essence. It is constructed around seven basic propositions, numbered by the natural numbers 1–7, with all other paragraphs in the text numbered by decimal expansions so that, e.g., paragraph 1.1 is (supposed to be) a further elaboration on proposition 1, 1.22 is an elaboration of 1.2, and so on.

The seven basic propositions are:

Ogden translation	Pears/McGuinness translation
1. The world is everything that is the case.	The world is all that is the case.
2. What is the case, the fact, is the existence of atomic facts.	What is the case—a fact—is the existence of states of affairs.
3. The logical picture of the facts is the thought.	A logical picture of facts is a thought.
4. The thought is the significant proposition.	A thought is a proposition with sense.
5. Propositions are truth-functions of elementary propositions.	A proposition is a truth-function of elementary propositions.
(An elementary proposition is a truth function of itself.)	(An elementary proposition is a truth function of itself.)
6. The general form of truth-function is $[\bar{p}, \bar{\xi}, N(\bar{\xi})]$.	The general form of a truth-function is $[\bar{p}, \bar{\xi}, N(\bar{\xi})]$.
This is the general form of proposition.	This is the general form of a proposition.
7. Whereof one cannot speak, thereof one must be silent.	What we cannot speak about we must pass over in silence.

Clearly, the book addresses the central problems of philosophy which deal with the world, thought and language, and presents a 'solution' (as Wittgenstein terms it) of these problems that is grounded in logic and in the nature of representation. The world is represented by thought, which is a proposition with sense, since they all—world, thought, and proposition

—share the same logical form. Hence, the thought and the proposition can be pictures of the facts.

Starting with a seeming metaphysics, Wittgenstein sees the world as consisting of facts (1), rather than the traditional, atomistic conception of a world made up of objects. Facts are existent states of affairs (2) and states of affairs, in turn, are combinations of objects. "Objects are simple" (*TLP* 2.02) but objects can fit together in various determinate ways. They may have various properties and may hold diverse relations to one another. Objects combine with one another according to their logical, internal properties. That is to say, an object's internal properties determine the possibilities of its combination with other objects; this is its logical form. Thus, states of affairs, being comprised of objects in combination, are inherently complex. The states of affairs which do exist could have been otherwise. This means that states of affairs are either actual (existent) or possible. It is the totality of states of affairs—actual and possible—that makes up the whole of reality. The world is precisely those states of affairs which do exist.

The move to thought, and thereafter to language, is perpetrated with the use of Wittgenstein's famous idea that thoughts, and propositions, are pictures—"the picture is a model of reality" (*TLP* 2.12). Pictures are made up of elements that together constitute the picture. Each element represents an object, and the combination of elements in the picture represents the combination of objects in a state of affairs. The logical structure of the picture, whether in thought or in language, is isomorphic with the logical structure of the state of affairs which it pictures. More subtle is Wittgenstein's insight that the possibility of this structure being shared by the picture (the thought, the proposition) and the state of affairs is the pictorial form. "*That* is how a picture is attached to reality; it reaches right out to it" (*TLP* 2.1511). This leads to an understanding of what the picture can picture; but also what it cannot—its own pictorial form.

While "the logical picture of the facts is the thought" (3), in the move to language Wittgenstein continues to investigate the possibilities of significance for propositions (4). Logical analysis, in the spirit of Frege and Russell, guides the work, with Wittgenstein using logical calculus to carry out the construction of his system. Explaining that "Only the proposition has sense; only in the context of a proposition has a name meaning" (*TLP* 3.3), he provides the reader with the two conditions for sensical language. First, the structure of the proposition must conform to the constraints of logical form, and second, the elements of the proposition must have reference (*bedeutung*). These conditions have far-reaching implications. The analysis must culminate with a name being a primitive symbol for a (simple) object. Moreover, logic itself gives us the structure and limits of what can be said at all.

"The general form of a proposition is: This is how things stand" (*TLP* 4.5) and every proposition is either true or false. This bi-polarity of propositions enables the composition of more complex propositions from atomic ones by using truth-functional operators (5). Wittgenstein supplies, in the *Tractatus*, a vivid presentation of Frege's logic in the form of what has become known as 'truth-tables'. This provides the means to go back and analyze all propositions into their atomic parts, since "every statement about complexes can be analyzed into a statement about their constituent parts, and into those propositions which completely describe the complexes" (*TLP* 2.0201). He delves even deeper by then providing the general form of a truth-function (6). This form, $[\bar{p}, \bar{\xi}, N(\bar{\xi})]$, makes use of one formal operation ($N(\bar{\xi})$) and one propositional variable ($\bar{p}$) to represent Wittgenstein's claim that any proposition "is the result of successive applications" of logical operations to elementary propositions.

Having developed this analysis of world-thought-language, and relying on the one general form of the proposition, Wittgenstein can now assert that all meaningful propositions are of equal value. Subsequently, he ends the

journey with the admonition concerning what can (or cannot) and what should (or should not) be said (7), leaving outside the realm of the sayable propositions of ethics, aesthetics, and metaphysics.

2.2 Sense and Nonsense

In the *Tractatus* Wittgenstein's logical construction of a philosophical system has a purpose—to find the limits of world, thought and language; in other words, to distinguish between sense and nonsense. "The book will … draw a limit to thinking, or rather—not to thinking, but to the expression of thoughts …. The limit can … only be drawn in language and what lies on the other side of the limit will be simply nonsense" (*TLP* Preface). The conditions for a proposition's having sense have been explored and seen to rest on the possibility of representation or picturing. Names must have a *bedeutung* (reference/meaning), but they can only do so in the context of a proposition which is held together by logical form. It follows that only factual states of affairs which can be pictured can be represented by meaningful propositions. This means that what can be said are only propositions of natural science and leaves out of the realm of sense a daunting number of statements which are made and used in language.

There are, first, the propositions of logic itself. These do not represent states of affairs, and the logical constants do not stand for objects. "My fundamental thought is that the logical constants do not represent. That the *logic* of the facts cannot be represented" (*TLP* 4.0312). This is not a happenstance thought; it is fundamental precisely because the limits of sense rest on logic. Tautologies and contradictions, the propositions of logic, are the limits of language and thought, and thereby the limits of the world. Obviously, then, they do not picture anything and do not, therefore, have sense. They are, in Wittgenstein's terms, senseless (*sinnlos*). Propositions which do have sense are bipolar; they range within the truth-

conditions drawn by the truth-tables. But the propositions of logic themselves are "not pictures of the reality … for the one allows *every* possible state of affairs, the other *none*" (*TLP* 4.462). Indeed, tautologies (and contradictions), being senseless, are recognized as true (or false) "in the symbol alone … and this fact contains in itself the whole philosophy of logic" (*TLP* 6.113).

The characteristic of being senseless applies not only to the propositions of logic but also to mathematics or the pictorial form itself of the pictures that do represent. These are, like tautologies and contradictions, literally sense-less, they have no sense.

Beyond, or aside from, senseless propositions Wittgenstein identifies another group of statements which cannot carry sense: the nonsensical (*unsinnig*) propositions. Nonsense, as opposed to senselessness, is encountered when a proposition is even more radically devoid of meaning, when it transcends the bounds of sense. Under the label of *unsinnig* can be found various propositions: "Socrates is identical", but also "1 is a number" and "there are objects". While some nonsensical propositions are blatantly so, others seem to be meaningful—and only analysis carried out in accordance with the picture theory can expose their nonsensicality. Since only what is "in" the world can be described, anything that is "higher" is excluded, including the notion of limit and the limit points themselves. Traditional metaphysics, and the propositions of ethics and aesthetics, which try to capture the world as a whole, are also excluded, as is the truth in solipsism, the very notion of a subject, for it is also not "in" the world but at its limit.

Wittgenstein does not, however, relegate all that is not inside the bounds of sense to oblivion. He makes a distinction between *saying* and *showing* which is made to do additional crucial work. "What can be shown cannot be said," that is, what cannot be formulated in sayable (sensical)

propositions can only be shown. This applies, for example, to the logical form of the world, the pictorial form, etc., which show themselves in the form of (contingent) propositions, in the symbolism, and in logical propositions. Even the unsayable (metaphysical, ethical, aesthetic) propositions of philosophy belong in this group—which Wittgenstein finally describes as "things that cannot be put into words. They make themselves manifest. They are what is mystical" (*TLP* 6.522).

2.3 The Nature of Philosophy

Accordingly, "the word 'philosophy' must mean something which stands above or below, but not beside the natural sciences" (*TLP* 4.111). Not surprisingly, then, "most of the propositions and questions to be found in philosophical works are not false but nonsensical" (*TLP* 4.003). Is, then, philosophy doomed to be nonsense (*unsinnig*), or, at best, senseless (*sinnlos*) when it does logic, but, in any case, meaningless? What is left for the philosopher to do, if traditional, or even revolutionary, propositions of metaphysics, epistemology, aesthetics, and ethics cannot be formulated in a sensical manner? The reply to these two questions is found in Wittgenstein's characterization of philosophy: philosophy is not a theory, or a doctrine, but rather an activity. It is an activity of clarification (of thoughts), and more so, of critique (of language). Described by Wittgenstein, it should be the philosopher's routine activity: to react or respond to the traditional philosophers' musings by showing them where they go wrong, using the tools provided by logical analysis. In other words, by showing them that (some of) their propositions are nonsense.

"All propositions are of equal value" (*TLP* 6.4)—that could also be the fundamental thought of the book. For it employs a measure of the value of propositions that is done by logic and the notion of limits. It is here, however, with the constraints on the value of propositions, that the tension in the *Tractatus* is most strongly felt. It becomes clear that the notions

used by the *Tractatus*—the logical-philosophical notions—do not belong to the world and hence cannot be used to express anything meaningful. Since language, thought and the world, are all isomorphic, any attempt to say in logic (i.e., in language) "this and this there is in the world, that there is not" is doomed to be a failure, since it would mean that logic has got outside the limits of the world, i.e. of itself. That is to say, the *Tractatus* has gone over its own limits, and stands in danger of being nonsensical.

The "solution" to this tension is found in Wittgenstein's final remarks, where he uses the metaphor of the ladder to express the function of the *Tractatus*. It is to be used in order to climb on it, in order to "see the world rightly"; but thereafter it must be recognized as nonsense and be thrown away. Hence: "whereof one cannot speak, thereof one must be silent" (7).

2.4 Interpretative Problems

The *Tractatus* is notorious for its interpretative difficulties. In the decades that have passed since its publication it has gone through several waves of general interpretations. Beyond exegetical and hermeneutical issues that revolve around particular sections (such as the world/reality distinction, the difference between representing and presenting, the Frege/Russell connection to Wittgenstein, or the influence on Wittgenstein by existentialist philosophy) there are a few fundamental, not unrelated, disagreements that inform the map of interpretation. These revolve around the realism of the *Tractatus*, the notion of nonsense and its role in reading the *Tractatus* itself, and the reading of the *Tractatus* as an ethical tract.

There are interpretations that see the *Tractatus* as espousing realism, i.e., as positing the independent existence of objects, states of affairs, and facts. That this realism is achieved via a linguistic turn is recognized by all (or most) interpreters, but this linguistic perspective does no damage to the basic realism that is seen to start off the *Tractatus* ("The world is all that is

the case") and to run throughout the text ("Objects form the substance of the world" (*TLP* 2.021)). Such realism is also taken to be manifested in the essential bi-polarity of propositions; likewise, a straightforward reading of the picturing relation posits objects there to be represented by signs. As against these readings, more linguistically oriented interpretations give conceptual priority to the symbolism. When "reality is compared with propositions" (*TLP* 4.05), it is the form of propositions which determines the shape of reality (and not the other way round). In any case, the issue of realism (vs. anti-realism) in the *Tractatus* must address the question of the limits of language and the more particular question of what there is (or is not) beyond language. Subsequently, interpreters of the *Tractatus* have moved on to questioning the very presence of metaphysics within the book and the status of the propositions of the book themselves.

'Nonsense' became the hinge of Wittgensteinian interpretative discussion during the last decade of the 20th century. Beyond the bounds of language lies nonsense—propositions which cannot picture *anything*—and Wittgenstein bans traditional metaphysics to that area. The quandary arises concerning the question of what it is that inhabits that realm of nonsense, since Wittgenstein does seem to be saying that there is something there to be shown (rather than said) and does, indeed, characterize it as the 'mystical'. The traditional readings of the *Tractatus* accepted, with varying degrees of discomfort, the existence of that which is unsayable, that which cannot be put into words, the nonsensical. More recent readings tend to take nonsense more seriously as exactly that—nonsense. This also entails taking seriously Wittgenstein's words in 6.54—his famous ladder metaphor—and throwing out the *Tractatus* itself, including the distinction between what can be said and what can only be shown. The *Tractatus*, on this stance, does not point at ineffable truths (of, e.g., metaphysics, ethics, aesthetics, etc.), but should lead us away from such temptations. An accompanying discussion must then also deal with how this can be

recognized, what this can possibly mean, and how it should be used, if at all.

This discussion is closely related to what has come to be called the ethical reading of the *Tractatus*. Such a reading is based, first, on the supposed discrepancy between Wittgenstein's construction of a world-language system, which takes up the bulk of the *Tractatus*, and several comments that are made about this construction in the Preface to the book, in its closing remarks, and in a letter he sent to his publisher, Ludwig von Ficker, before publication. In these places, all of which can be viewed as external to the content of the *Tractatus*, Wittgenstein preaches silence as regards anything that is of importance, including the 'internal' parts of the book which contain, in his own words, "the final solution of the problems [of philosophy]." It is the importance given to the ineffable that can be viewed as an ethical position. "My work consists of two parts, the one presented here plus all that I have *not* written. And it is precisely this second part that is the important point. For the ethical gets its limit drawn from the inside, as it were, by my book; … I've managed in my book to put everything firmly into place by being silent about it …. For now I would recommend you to read the *preface* and the *conclusion*, because they contain the most direct expression of the point" (*ProtoTractatus*, p.16). Obviously, such seemingly contradictory tensions within and about a text—written by its author—give rise to interpretative conundrums.

There is another issue often debated by interpreters of Wittgenstein, which arises out of the questions above. This has to do with the continuity between the thought of the early and later Wittgenstein. Again, the 'standard' interpretations were originally united in perceiving a clear break between the two distinct stages of Wittgenstein's thought, even when ascertaining some developmental continuity between them. And again, the more recent interpretations challenge this standard, emphasizing that the

fundamental therapeutic motivation clearly found in the later Wittgenstein should also be attributed to the early.

3. The Later Wittgenstein

3.1 Transition and Critique of *Tractatus*

The idea that philosophy is not a doctrine, and hence should not be approached dogmatically, is one of the most important insights of the *Tractatus*. Yet, as early as 1931, Wittgenstein referred to his own early work as dogmatic ("On Dogmatism" in *VC*, p. 182. Wittgenstein used this term to designate any conception which allows for a gap between question and answer, such that the answer to the question could be found at a later date. The complex edifice of the *Tractatus* is built on the assumption that the task of logical analysis was to discover the elementary propositions, whose form was not yet known. What marks the transition from early to later Wittgenstein can be summed up as the *total* rejection of dogmatism, i.e., as the working out of *all* the consequences of this rejection. The move from the realm of logic to that of ordinary language as the center of the philosopher's attention; from an emphasis on definition and analysis to 'family resemblance' and 'language-games'; and from systematic philosophical writing to an aphoristic style—all have to do with this transition towards anti-dogmatism in its extreme. It is in the *Philosophical Investigations* that the working out of the transitions comes to culmination. Other writings of the same period, though, manifest the same anti-dogmatic stance, as it is applied, e.g., to the philosophy of mathematics or to philosophical psychology.

3.2 *Philosophical Investigations*

Philosophical Investigations was published posthumously in 1953. It was edited by G. E. M. Anscombe and Rush Rhees and translated by

Anscombe. It comprised two parts. Part I, consisting of 693 numbered paragraphs, was ready for printing in 1946, but rescinded from the publisher by Wittgenstein. Part II was added on by the editors, trustees of his *Nachlass*. In 2009 a new edited translation, by P. M. S. Hacker and Joachim Schulte, was published; Part II of the earlier translation was here labeled "Philosophy of Psychology – A Fragment" (PPF).

In the Preface to *PI*, Wittgenstein states that his new thoughts would be better understood by contrast with and against the background of his old thoughts, those in the *Tractatus*; and indeed, most of Part I of *PI* is essentially critical. Its new insights can be understood as primarily exposing fallacies in the traditional way of thinking about language, truth, thought, intentionality, and, perhaps mainly, philosophy. In this sense, it is conceived of as a *therapeutic* work, viewing philosophy itself as *therapy*. (Part II (PPF), focusing on philosophical psychology, perception etc., was not as critical. Rather, it pointed to new perspectives (which, undoubtedly, are not disconnected from the earlier critique) in addressing specific philosophical issues. It is, therefore, more easily read alongside Wittgenstein's other writings of the later period.)

PI begins with a quote from Augustine's *Confessions* which "give us a particular picture of the essence of human language," based on the idea that "the words in language name objects," and that "sentences are combinations of such names" (*PI* 1). This picture of language cannot be relied on as a basis for metaphysical, epistemic or linguistic speculation. Despite its plausibility, this reduction of language to representation cannot do justice to the whole of human language; and even if it is to be considered a picture of only the representative function of human language, it is, as such, a poor picture. Furthermore, this picture of language is at the base of the whole of traditional philosophy, but, for Wittgenstein, it is to be shunned in favor of a new way of looking at both language and philosophy. The *Philosophical Investigations* proceeds to

offer the new way of looking at language, which will yield the view of philosophy as therapy.

3.3 Meaning as Use

"For a *large* class of cases of the employment of the word 'meaning'—though not for all—this word can be explained in this way: the meaning of a word is its use in the language" (*PI* 43). This basic statement is what underlies the change of perspective most typical of the later phase of Wittgenstein's thought: a change from a conception of meaning as representation to a view which looks to use as the crux of the investigation. Traditional theories of meaning in the history of philosophy were intent on pointing to something exterior to the proposition which endows it with sense. This 'something' could generally be located either in an objective space, or inside the mind as mental representation. As early as 1933 (*The Blue Book*) Wittgenstein took pains to challenge these conceptions, arriving at the insight that "if we had to name anything which is the life of the sign, we should have to say that it was its *use*" (*BB* 4). Ascertainment of the use (of a word, of a proposition), however, is not given to any sort of constructive theory building, as in the *Tractatus*. Rather, when investigating meaning, the philosopher must "look and see" the variety of uses to which the word is put. An analogy with tools sheds light on the nature of words. When we think of tools in a toolbox, we do not fail to see their variety; but the "functions of words are as diverse as the functions of these objects" (*PI* 11). We are misled by the uniform appearance of our words into theorizing upon meaning: "Especially when we are doing philosophy!" (*PI* 12)

So different is this new perspective that Wittgenstein repeats: "Don't think, but look!" (*PI* 66); and such looking is done *vis a vis* particular cases, not generalizations. In giving the meaning of a word, any explanatory generalization should be replaced by a description of use. The

traditional idea that a proposition houses a content and has a restricted number of Fregean forces (such as assertion, question and command), gives way to an emphasis on the diversity of uses. In order to address the countless multiplicity of uses, their un-fixedness, and their being part of an activity, Wittgenstein introduces the key concept of 'language-game'. He never explicitly defines it since, as opposed to the earlier 'picture', for instance, this new concept is made to do work for a more fluid, more diversified, and more activity-oriented perspective on language.

3.4 Language-games and Family Resemblance

Throughout the *Philosophical Investigations*, Wittgenstein returns, again and again, to the concept of language-games to make clear his lines of thought concerning language. Primitive language-games are scrutinized for the insights they afford on this or that characteristic of language. Thus, the builders' language-game (*PI* 2), in which a builder and his assistant use exactly four terms (block, pillar, slab, beam), is utilized to illustrate that part of the Augustinian picture of language which might be correct but which is, nevertheless, strictly limited. 'Regular' language-games, such as the astonishing list provided in *PI* 23 (which includes, e.g., reporting an event, speculating about an event, forming and testing a hypothesis, making up a story, reading it, play-acting, singing catches, guessing riddles, making a joke, translating, asking, thanking, and so on), bring out the openness of our possibilities in using language and in describing it.

Language-games are, first, a part of a broader context termed by Wittgenstein a form of life (see below). Secondly, the concept of language-games points at the rule-governed character of language. This does not entail strict and definite systems of rules for each and every language-game, but points to the conventional nature of this sort of human activity. Still, just as we cannot give a final, essential definition of 'game',

so we cannot find "what is common to all these activities and what makes them into language or parts of language" (*PI* 65).

It is here that Wittgenstein's rejection of general explanations, and definitions based on sufficient and necessary conditions, is best pronounced. Instead of these symptoms of the philosopher's "craving for generality", he points to 'family resemblance' as the more suitable analogy for the means of connecting particular uses of the same word. There is no reason to look, as we have done traditionally—and dogmatically—for one, essential core in which the meaning of a word is located and which is, therefore, common to all uses of that word. We should, instead, travel with the word's uses through "a complicated network of similarities overlapping and criss-crossing" (*PI* 66). Family resemblance also serves to exhibit the lack of boundaries and the distance from exactness that characterize different uses of the same concept. Such boundaries and exactness are the definitive traits of form—be it Platonic form, Aristotelian form, or the general form of a proposition adumbrated in the *Tractatus*. It is from such forms that applications of concepts can be deduced, but this is precisely what Wittgenstein now eschews in favor of appeal to similarity of a kind with family resemblance.

3.5 Rule-following and Private Language

One of the issues most associated with the later Wittgenstein is that of rule-following. Rising out of the considerations above, it becomes another central point of discussion in the question of what it is that can apply to all the uses of a word. The same dogmatic stance as before has it that a rule is an abstract entity—transcending all of its particular applications; knowing the rule involves grasping that abstract entity and thereby knowing how to use it.

Wittgenstein begins his exposition by introducing an example: "… we get [a] pupil to continue a series (say '+ 2') beyond 1000—and he writes 1000, 1004, 1008, 1012 (*PI* 185)". What do we do, and what does it mean, when the student, upon being corrected, answers "But I did go on in the same way"? Wittgenstein proceeds (mainly in *PI* 185–243, but also elsewhere) to dismantle the cluster of attendant questions: How do we learn rules? How do we follow them? Wherefrom the standards which decide if a rule is followed correctly? Are they in the mind, along with a mental representation of the rule? Do we appeal to intuition in their application? Are they socially and publicly taught and enforced? In typical Wittgensteinian fashion, the answers are not pursued positively; rather, the very formulation of the questions as legitimate questions with coherent content is put to the test. For indeed, it is both the Platonistic and mentalistic pictures which underlie asking questions of this type, and Wittgenstein is intent on freeing us from these assumptions. Such liberation involves elimination of the need to posit any sort of external or internal authority beyond the actual applications of the rule.

These considerations lead to *PI* 201, often considered the climax of the issue: "This was our paradox: no course of action could be determined by a rule, because every course of action can be made out to accord with the rule. The answer was: if everything can be made out to accord with the rule, then it can also be made out to conflict with it. And so there would be neither accord nor conflict here." Wittgenstein's formulation of the problem, now at the point of being a "paradox", has given rise to a wealth of interpretation and debate since it is clear to all that this is the crux of the general issue of meaning, and of understanding and using a language. One of the influential readings of the problem of following a rule (introduced by Fogelin 1976 and Kripke 1982) has been the interpretation, according to which Wittgenstein is here voicing a skeptical paradox and offering a skeptical solution. That is to say, there are no facts that determine what counts as following a rule, no real grounds for saying that someone is

indeed following a rule, and Wittgenstein accepts this skeptical challenge (by suggesting other conditions that might warrant our asserting that someone is following a rule). This reading has been challenged, in turn, by several interpretations (such as Baker and Hacker 1984, McGinn1984, and Cavell 1990), while others have provided additional, fresh perspectives (e.g., Diamond, "Rules: Looking in the Right Place" in Phillips and Winch 1989, and several in Miller and Wright 2002).

Directly following the rule-following sections in *PI*, and therefore easily thought to be the upshot of the discussion, are those sections called by interpreters "the private-language argument". Whether it be a veritable argument or not (and Wittgenstein never labeled it as such), these sections point out that for an utterance to be meaningful it must be possible in principle to subject it to public standards and criteria of correctness. For this reason, a private-language, in which "words … are to refer to what only the speaker can know—to his immediate private sensations …" (*PI* 243), is not a genuine, meaningful, rule-governed language. The signs in language can only function when there is a possibility of judging the correctness of their use, "so the use of [a] word stands in need of a justification which everybody understands" (*PI* 261).

3.6 Grammar and Form of Life

Grammar, usually taken to consist of the rules of correct syntactic and semantic usage, becomes, in Wittgenstein's hands, the wider—and more elusive—network of rules which determine what linguistic move is allowed as making sense, and what isn't. This notion replaces the stricter and purer logic, which played such an essential role in the *Tractatus* in providing a scaffolding for language and the world. Indeed, "*Essence* is expressed in grammar … Grammar tells what kind of object anything is. (Theology as grammar)" (*PI* 371, 373). The "rules" of grammar are not mere technical instructions from on-high for correct usage; rather, they

express the norms for meaningful language. Contrary to empirical statements, rules of grammar describe how we use words in order to both justify and criticize our particular utterances. But as opposed to grammar-book rules, they are not idealized as an external system to be conformed to. Moreover, they are not appealed to explicitly in any formulation, but are used in cases of philosophical perplexity to clarify where language misleads us into false illusions. Thus, "I can know what someone else is thinking, not what I am thinking. It is correct to say 'I know what you are thinking', and wrong to say 'I know what I am thinking.' (A whole cloud of philosophy condensed into a drop of grammar.)" (*Philosophical Investigations* 1953, p.222).

Grammar is not abstract, it is situated within the regular activity with which language-games are interwoven: "… the word 'language-*game*' is used here to emphasize the fact that the *speaking* of language is part of an activity, or of a form of life" (*PI* 23). What enables language to function and therefore must be accepted as "given" are precisely forms of life. In Wittgenstein's terms, "It is not only agreement in definitions but also (odd as it may sound) in judgments that is required" (*PI* 242), and this is "agreement not in opinions, but rather in form of life" (*PI* 241). Used by Wittgenstein sparingly—five times in the *Investigations*—this concept has given rise to interpretative quandaries and subsequent contradictory readings. Forms of life can be understood as changing and contingent, dependent on culture, context, history, etc; this appeal to forms of life grounds a relativistic reading of Wittgenstein. On the other hand, it is the form of life common to humankind, "shared human behavior" which is "the system of reference by means of which we interpret an unknown language" (*PI* 206). This might be seen as a universalistic turn, recognizing that the use of language is made possible by the human form of life.

3.7 The Nature of Philosophy

In his later writings Wittgenstein holds, as he did in the *Tractatus*, that philosophers do not—or should not—supply a theory, neither do they provide explanations. "Philosophy just puts everything before us, and neither explains nor deduces anything.—Since everything lies open to view there is nothing to explain" (*PI* 126). The anti-theoretical stance is reminiscent of the early Wittgenstein, but there are manifest differences. Although the *Tractatus* precludes philosophical theories, it does construct a systematic edifice which results in the general form of the proposition, all the while relying on strict formal logic; the *Investigations* points out the therapeutic non-dogmatic nature of philosophy, verily instructing philosophers in the ways of therapy. "The work of the philosopher consists in marshalling reminders for a particular purpose" (*PI* 127). Working with reminders and series of examples, different problems are solved. Unlike the *Tractatus* which advanced one philosophical method, in the *Investigations* "there is not a single philosophical method, though there are indeed methods, different therapies, as it were" (*PI* 133d). This is directly related to Wittgenstein's eschewal of *the* logical form or of any a-priori generalization that can be discovered or made in philosophy. Trying to advance such general theses is a temptation which lures philosophers; but the real task of philosophy is both to make us aware of the temptation and to show us how to overcome it. Consequently "a philosophical problem has the form: 'I don't know my way about.'" (*PI* 123), and hence the aim of philosophy is "to show the fly the way out of the fly-bottle" (*PI* 309).

The style of the *Investigations* is strikingly different from that of the *Tractatus*. Instead of strictly numbered sections which are organized hierarchically in programmatic order, the *Investigations* fragmentarily voices aphorisms about language-games, family resemblance, forms of life, "sometimes jumping, in a sudden change, from one area to another" (*PI* Preface). This variation in style is of course essential and is "connected with the very nature of the investigation" (*PI* Preface). As a matter of fact, Wittgenstein was acutely aware of the contrast between the

two stages of his thought, suggesting publication of both texts together in order to make the contrast obvious and clear.

Still, it is precisely via the subject of the nature of philosophy that the fundamental continuity between these two stages, rather than the discrepancy between them, is to be found. In both cases philosophy serves, first, as critique of language. It is through analyzing language's illusive power that the philosopher can expose the traps of meaningless philosophical formulations. This means that what was formerly thought of as a philosophical problem may now dissolve "and this simply means that the philosophical problems should *completely* disappear" (*PI* 133). Two implications of this diagnosis, easily traced back in the *Tractatus*, are to be recognized. One is the inherent dialogical character of philosophy, which is a responsive activity: difficulties and torments are encountered which are then to be dissipated by philosophical therapy. In the *Tractatus*, this took the shape of advice: "The correct method in philosophy would really be the following: to say nothing except what can be said, i.e. propositions of natural science … and then whenever someone else wanted to say something metaphysical, to demonstrate to him that he had failed to give a meaning to certain signs in his propositions" (*TLP* 6.53) The second, more far- reaching, "discovery" in the *Investigations* "is the one that enables me to break off philosophizing when I want to" (*PI* 133). This has been taken to revert back to the ladder metaphor and the injunction to silence in the *Tractatus*.

3.8 After the *Investigations*

It has been submitted that the writings of the period from 1946 until his death (1951) constitute a distinctive phase of Wittgenstein's thought. These writings include, in addition to the second part of the first edition of the *Philosophical Investigations*, texts edited and collected in volumes such as *Remarks on Colour*, *Remarks on the Philosophy of Psychology*,

Zettel, *On Certainty*, and parts of *The Foundations of Mathematics*. Besides dealing with mathematics and psychology, this is the stage at which Wittgenstein most seriously pursued questions traditionally recognized as epistemological. *On Certainty* tackles skeptical doubts and foundational solutions but is, in typical Wittgensteinian fashion, a work of therapy which discounts presuppositions common to both. This is intimately related to another of *On Certainty*'s themes—the primacy of the deed to the word, or, in Wittgenstein's *PI* terminology, of form of life to grammar. The general tenor of all the writings of this last period can thence be viewed as, on the one hand, a move away from the critical (some would say destructive) positions of the *Investigations* to a more positive perspective on the same problems that had been facing him since his early writings; on the other hand, this move does not constitute a break from the later period but is more properly viewed as its continuation, in a new light.

Bibliography

Wittgenstein's Works

The Big Typescript: TS 213, German English Scholars' Edition, 2005, C. Grant Luckhardt and Maximilian E. Aue (trans.), Oxford: Wiley-Blackwell.

The Blue and Brown Books (BB), 1958, Oxford: Blackwell.

Culture and Value, 1980, G. H. von Wright (ed.), P. Winch (trans.), Oxford: Blackwell.

Last Writings on the Philosophy of Psychology, vol. 1, 1982, vol. 2, 1992, G. H. von Wright and H. Nyman (eds.), trans. C. G. Luckhardt and M. A. E. Aue (trans.), Oxford: Blackwell.

"A Lecture on Ethics", 1965, *The Philosophical Review*, 74: 3–12.

Lectures and Conversations on Aesthetics, Psychology and Religious Belief, 1966, C. Barrett (ed.), Oxford: Blackwell.

Logical Empiricism

First published Mon Apr 4, 2011; substantive revision Wed Apr 5, 2017

Logical empiricism is a philosophic movement rather than a set of doctrines, and it flourished in the 1920s and 30s in several centers in Europe and in the 40s and 50s in the United States. It had several different leaders whose views changed considerably over time. Moreover, these thinkers differed from one another, often sharply. Because logical empiricism is here construed as a movement rather than as doctrine, there is probably no important position that all logical empiricists shared—including, surprisingly enough, empiricism. And while most participants in the movement were empiricists of one form or another, they disagreed on what the best form of empiricism was and on the cognitive status of empiricism. What held the group together was a common concern for scientific methodology and the important role that science could play in reshaping society. Within that scientific methodology the logical empiricists wanted to find a natural and important role for logic and mathematics and to find an understanding of philosophy according to which it was part of the scientific enterprise.

The following discussion of logical empiricism is organized under five headings:

- 1. Mapping the Movement
- 2. Background
- 3. Some Major Participants in the Movement
- 4. Issues
    - 4.1 Empiricism, Verificationism, and Anti-metaphysics
    - 4.2 Analyticity
    - 4.3 Unity of Science and Reduction
    - 4.4 Probability

1. Mapping the Movement

The term 'logical empiricism' has no very precise boundaries and still less that distinguishes it from 'logical positivism'. It is therefore hard to map. 'Logical empiricism' here includes three groups: (1) the Vienna Circle, here taken broadly to include those who were part of various private discussion groups, especially that around Moritz Schlick, and also the members of the more public Ernst Mach Society (Verein Ernst Mach), (2) the smaller, but perhaps more influential Berlin Society for Empirical Philosophy (later called the Berlin Society for Scientific Philosophy), and (3) those influenced by or who interacted with members of the first two groups and shared an intellectual kinship with them. Besides Vienna and Berlin, there were important centers of the movement in England, France, Scandinavia, at several universities in the U.S., and even China. This characterization includes thinkers who disagreed with doctrines espoused by members of the original groups and even some who defined themselves in opposition to the movement. This results in a vague boundary, but it suffices to identify a movement in which a large number of able philosophers self-consciously participated and to distinguish logical empiricism from other movements.

It does not, however, distinguish logical empiricism from logical positivism, and it is doubtful that any principled such boundary can be drawn along doctrinal or sociological lines (Uebel 2013). Usually when distinctions are drawn, 'logical empiricism' is the wider term. Members of

the Berlin group never used the term 'positivism' about themselves, but did use it concerning some unnamed Viennese in stressing their differences from the latter. In any case, these differences, even if real, were smaller than the differences within the Vienna Circle on one hand or within the Berlin group on the other. 'Positivist' is a term usually applied by opponents of various doctrines. It was used by some of the Viennese logical empiricists about themselves but generally with caution and in stressing the differences between their own views and those of the 19th century positivists. The one philosopher who would have unhesitatingly described himself as (having been) a logical positivist was A.J. Ayer.

Another way of mapping the boundaries of logical empiricism is to list the specific philosophers who were centrally or peripherally part of it. This included many of the most important philosophers of the mid-twentieth century. Hans Hahn, Moritz Schlick, Rudolf Carnap, and Otto Neurath were leaders of the Vienna Circle, and Kurt Gödel regularly attended its meetings. The list of its members, visitors, and interlocutors is staggering, including A.J. Ayer, Herbert Feigl, Philipp Frank, Hans Hahn, Carl Hempel, Karl Menger, Richard von Mises, Ernest Nagel, Karl Popper, W.V. Quine, Frank Ramsay, Hans Reichenbach, Alfred Tarski, Friedrich Waismann, and Ludwig Wittgenstein, among many others. Not all of these would admit to being part of the logical empiricist movement, of course, but a case can be made that all contributed to it. The Berlin Society for Empirical (or Scientific) Philosophy was, as stated, smaller but perhaps more influential. Led by Hans Reichenbach, it included Kurt Grelling, Walter Dubislav, Kurt Lewin, Richard von Mises, Paul Oppenheim, and others. Hempel took his doctorate in Berlin, working with Reichenbach until the latter was forced to leave in 1933. Hempel also spent time in Vienna and Prague. Of course, among the foremost associates of the Berlin Society was Albert Einstein, who was also in Berlin also until 1933.

There was also an important group of logicians in Warsaw of which Alfred Tarski is the best known. Tarski interacted significantly with the logical empiricists in Vienna, Berlin, and the U.S., but it is more reasonable to classify the Polish logicians as an allied group rather than include them within the logical empiricist movement.

Because of the catastrophic dislocations of Europe in the 1930s, the main focus of the logical empiricism moved from central Europe to America by the close of that decade. *Erkenntnis*, the main journal of the movement, which had been edited by Reichenbach and Carnap, ceased publication by 1940. In 1930 Feigl moved to the U.S., and Carnap moved to Chicago in 1936. Hempel came to Chicago and Menger to Notre Dame in 1937. The ensuing years witnessed a massive exodus to America from central Europe. Reichenbach arrived in the U.S. in 1938 after five years in Turkey. Also in 1938 Gustav Bergmann and Philipp Frank emigrated. Edgar Zilsel came in 1939. Alfred Tarski was on a visit to the U.S. when Poland was invaded in 1939, and so he stayed. And by 1940 Richard von Mises was also in America.

In the U.S., these exiles were joined by the Americans Nelson Goodman, Charles Morris, W.V. Quine, Ernest Nagel, and, after the war, by Reichenbach's UCLA students Hilary Putnam and Wesley Salmon. Adolf Grünbaum can also be considered as clearly in the Reichenbach lineage. And Wilfrid Sellars was, in his early years, a close associate of Feigl. The American incarnation of the logical empiricist movement enjoyed generally good relations with the American pragmatists, not only because many of the logical empiricists had a strong pragmatist component to their philosophy, but also because the pragmatists and logical empiricists shared a common concern for empirical methodology in the service of social reform. Institutionally, the movement was represented in most major American universities, and such journals as *Philosophy of Science* (with Carnap and Feigl on the Editorial Board and Reichenbach and Schlick on

the Advisory Board) and *Philosophical Studies* (founded and edited for many years by Feigl and Sellars) provided ample outlet for their publications. In addition, the Inter-Scientific Discussion Group was founded by Philipp Frank at Harvard. That grew into the Institute for the Unity of Science, called by some the Vienna Circle in exile. Meanwhile in Chicago the *Encyclopedia of Unified Science* was established with Neurath, Carnap, and Morris as its editors.

But even from late 30s onward the movement was hardly limited to America. Ayer remained in England. Wittgenstein returned to Cambridge in 1929, but with regular visits to Vienna, including those on which he discussed issues surrounding a strong version of verificationism with Schlick and Waismann. Popper fled to New Zealand in 1937, and in 1946 moved to the London School of Economics. Neurath fled from Vienna to the Hague and then again in 1940 to England, where he remained till his death in 1945. Friedrich Waismann went to England in 1937. In 1939 Rose Rand, a less well-known member of the Vienna Circle, fled to England and then in 1954 emigrated once more to the U.S. There were like-minded thinkers in Scandinavia (such as Jørgen Jørgensen, Eino Kaila, and Arne Naess) and as far away as Argentina (H.A. Lindemann) and China (Tscha Hung).

It is impossible to say when logical empiricism ceased to be sufficiently cohesive to be identifiable as a continuing movement. Certainly by 1960 a great many philosophers, including many who had earlier clearly been part of the movement, were identifying themselves in opposition to what they took to be logical empiricism. And some members simply changed their minds or pursued different projects. Logical empiricism probably never commanded the assent of the majority of philosophers in either Europe or America, and by 1970 the movement was pretty clearly over—though with lasting influence whether recognized or not. In the 1980s there was a resurgence of historical interest in logical empiricism. That historical

interest continues to clear away many of the caricatures and misconceptions about the logical empiricists. Among the major results of this work is the recognition of the tremendous variety and subtlety of views represented within the movement and the fact that many of the arguments later deployed by critics of logical empiricism had been pioneered by the logical empiricists themselves.

Given the emphasis on science and its technical apparatus, social renewal, clarity and rationality of belief, functionality, and above all the palpable sense of doing philosophy in an importantly new way, it is reasonable to associate logical empiricism with other forms of European modernism in the 1920s and 30s, such as Neue Sachlichkeit in art and the Bauhaus in architecture and design, and with mid-century modernism as well as with political liberalism, from the New Deal to the Great Society in the United States. There have been recognizably modernist developments in various fields including philosophy for centuries.

2. Background

With a movement as large and complex as logical empiricism a great many factors went into raising the questions it would address, making them seem urgent, and making it seem as though the intellectual resources it would need to address these questions were either at hand or could be developed.

One long-term process with profound implications was the steady departure of the various sciences from philosophy to form autonomous disciplines. By early in the twentieth century mathematics, physics, chemistry, biology, and the social sciences were all pursued professionally and independently from philosophy. And psychology was just separating from philosophy. Yes, there were polymaths who could and did pursue a science and philosophy professionally. Those were increasingly rare,

though single-discipline scientists did from time to time make philosophic-seeming pronouncements. But they did so from outside the field. This pattern of steady departures raised the pressing question: What sort of thing remained behind? Once mathematics and the empirical sciences all left, what was left for philosophy?

The nature of philosophy was always a vexed philosophic question, but now it was particularly insistent. Surely there was no domain of empirical facts that philosophy could call its own. All that real estate had been parceled out. One answer available at the time that logical empiricism flourished was that the genuinely philosophic remainder after the departure of the sciences is somehow deeper than the empirical sciences and gets at matters, perhaps cultural ones, that are more profound and important than anything that empirical science even can address. This is either because on this conception philosophy has a mode of access or "evidence" that the empirical sciences do not and cannot have, or because the very idea of fidelity to evidence and punctilious argument is somehow small-minded.

The logical empiricists found this answer unappealing. Indeed, this conception of philosophy is precisely what Carnap means by 'metaphysics'. (As a consequence, what Carnap meant by that word is different from what late twentieth and early twenty-first century philosophers generally mean in describing their own work as analytic metaphysics.) The logical empiricists were eager to conceive of their enterprise as scientific and to engage in philosophy only insofar as it was also scientific. This science need not be empirical and need not include all that was traditional in philosophy that had not been incorporated into the independent sciences. The decision to be scientific can hardly be the end of the story. It requires rather better and more detailed answers to questions about what scientific methods are, how the mathematical (and

other apparently non-empirical sciences) fit together with the empirical ones, and what, more precisely, philosophy's role was.

A second series of developments that raised questions for logical empiricism to address were developments in the sciences themselves, especially the rise of non-Euclidean geometries in mathematics and the establishment of relativity theory in physics. These posed a serious challenge to what would otherwise be an attractive scientific philosophy, namely some version of Kantianism. Kant had recognized that the best of modern science was often mathematical in character and had labored to integrate both geometry and arithmetic into our empirical picture of the world. He had held that we could not represent the world except as a Euclidean structure and hence Euclidean geometry was, a priori, a permanent feature of any future physics. The demonstration that non-Euclidean pure geometrical structures were as consistent as Euclidean ones and that spaces can indeed be represented as a non-Euclidean manifolds was one half of the problem. The other half came when Einstein argued convincingly that physical space was best described as a non-Euclidean manifold of non-constant curvature. Plainly Euclidean geometry could not be guaranteed a future physics. Modern mathematical logic also posed a problem for other Kantian claims, but not in the same wrenching way.

Many logical empiricists started out as neo-Kantians: Reichenbach, Carnap, Schlick, and even Hempel (until he studied with Reichenbach, who by that time had revised his view). The difficulties with geometry and relativity certainly do not refute all forms of neo-Kantianism, but the difficulties are quite real nonetheless. The need is to understand how mathematics can be integrated into what is otherwise an empirical enterprise, i.e., physics, chemistry, biology, etc. Addressing this need was to be a major part of the logical empiricist program.

The background of logical empiricism described so far has been confined to the academic world, but events outside that domain shaped the movement as well. World War I was an unmitigated disaster for central Europe, followed by economic turmoil in the 20s and political upheavals of the 30s. It is hard to exaggerate these changes. Monarchies that had stood for centuries disappeared overnight and their empires disintegrated. This level of political convulsion had not been seen since the French Revolution, and that earlier upheaval was comparatively confined. Cultural changes were equally profound, and these were reflected by radical departures in the arts such as painting, music, and architecture, and even more importantly in new modes of living.

The logical empiricists were no mere bystanders. They, or at least the main leaders of the movement, were politically and culturally engaged. Even more important, this engagement was accompanied by the conviction that their cultures were incapable of the necessary reform and renewal because people were in effect enslaved by unscientific, metaphysical ways of thinking. Such ways of thinking might be exemplified in theology, in the racial hatreds of the day, in conceptions of property, and in traditional ideas about the "proper" roles of men and women in society. So to articulate a "scientific world conception" and to defend it against metaphysics was not just to express an academic position in the narrow sense. It was a political act as well; it was to strike a blow for the liberation of the mind. To articulate scientific methods and a scientific conception of philosophy was the essential first step in the reform of society and in the emancipation of humankind (Carnap 1958/2017, Creath 2009, Uebel 2012.

If all of this sounds like something out of the 18$^{\text{th}}$ century Enlightenment, the analogy was not lost on the logical empiricists themselves. André Carus has argued that this is exactly what Carnap had in mind by "explication" (Carus 2007). Neurath frequently drew parallels between the

logical empiricists' anti-metaphysical program and the earlier Enlightenment ambitions. Certainly Kant had inveighed against the metaphysics of his time, and the anti-metaphysical tradition remained strong within the scientific community through the 19th century.

The point so far was not to ask whether the logical empiricists were right in any of this. That question will come up later. So far the issue has been only to see the motivations that the logical empiricists had—and from their point of view—for addressing certain questions and for thinking that answers to those questions were urgently needed. None of this, however, says why the logical empiricists thought they had or could have the means to answer these questions. To that we now turn.

Since Newton the most paradigmatic examples of empirical science were those claims, usually quantitative ones, that were properly inferred from or appropriately confirmed by experience. Speaking very informally, these are the ones that we have good reason to believe or at least better reason to believe than the available alternatives. The problem, of course, is to specify the form of proper inferences, the form of an appropriate confirmation relation, and/or the structure of good reasons. The task is daunting, but logic in a suitably broad sense seems to be the right tool. Still speaking informally, logic seems to give us the structure of (good) reasoning. There are other conceptions of logic, of course, but this is a standard one and pretty well describes what the movement needed.

If logic was the tool that was wanted, it was newly ready for service. The progress of modern mathematical logic from Bolzano through Russell and beyond was truly impressive. Arguably, it could now express all parts of classical mathematics. Besides the first order predicate calculus one would need either set theory or higher order logic, but these were recent developments as well. Logic, like the empirical sciences, was progressive and could be approached cooperatively by more than one investigator. In

Our Knowledge of the External World (1914) Russell had even positioned logic as *the* locus of scientific method in philosophy. It is small wonder then that those who were looking for something scientific in what was left of philosophy turned to logic. Wittgenstein's no-content theory of logic in the *Tractatus* (1921/1922) was tantalizingly suggestive about how mathematics could be integrated into an overall empirical theory of the world. Wittgenstein also expressed a radical verificationism in the early 1930s in his conversations with Schlick, Waismann, and other members of the Vienna Circle. Many of the logical empiricists in turn could see in some version of that verificationism the ideal tool with which to carry out their anti-metaphysical program. There was, naturally, much left to accomplish, but even with Gödel's results one could expect that further impressive strides in logic could be made. Indeed, much was accomplished even if the perfect account of scientific reasoning proved elusive. Perfection is elusive in all the sciences, but that is no reason for despair.

3. Some Major Participants in the Movement

The logical empiricist movement is the sum of the interwoven trajectories of its members, so one way of describing that movement is to trace those various trajectories. To do so in detail for all those involved would take rather longer than the movement lasted. That would be inappropriate for one entry in an encyclopedia, especially one in which entries for many of the members will appear independently. The thumbnail sketches of the work of some representative figures below show the breadth and international character of the movement. While the list is long, it covers only a small fraction of those involved and leaves out many important thinkers.

A.J. Ayer (1910–1989)

An English philosopher in the tradition of British empiricism, Ayer

visited the Vienna Circle in 1932–33. His book *Language, Truth, and Logic* (1936) was a best seller after World War II and represents logical positivism to many English speakers

Rudolf Carnap (1891–1970)

German by birth, he taught in Vienna, Prague, Chicago, and Los Angeles. He was one of the leaders of the Vienna Circle and of logical empiricism, especially of those within the movement whose formulations were more liberal, e.g., with respect to the criterion of verification. He defended logical and methodological pluralism and worked to develop an epistemic approach to probability.

Walter Dubislav (1895–1937)

A German logician and philosopher of science, Dubislav was one of the founders, with Reichenbach and Grelling, of the Berlin Society of Empirical (later Scientific) Philosophy.

Herbert Feigl (1902–1988)

Born in what is now the Czech Republic, Feigl studied in Vienna with Schlick and Hahn. He emigrated to the U.S. before most other logical empiricists would do so. He taught at the Universities of Iowa and Minnesota and founded both *Philosophical Studies*, with Wilfrid Sellars, and the Minnesota Center for the Philosophy of Science. He is best known for his work on the mind-body problem.

Philipp Frank (1884–1966)

This Viennese physicist and philosopher of science taught at Vienna, Prague, and Harvard. He was part of a discussion group with Hahn, Neurath, and others that preceded the Vienna Circle. At Harvard he founded the Inter-Scientific Discussion Group that developed into the Institute for the Unity of Science. He was also one of the founders of the Boston Colloquium in the Philosophy of Science.

Kurt Gödel (1906–1978)

Born in what is now Slovakia, Gödel took his doctorate under Hahn in Vienna, studying with Carnap and Schlick as well. He also

regularly attended Vienna Circle meetings and taught in Vienna. The bulk of his career was spent at the Institute for Advanced Study at Princeton. He is best known for his spectacular incompleteness theorems, and his Platonist orientation toward mathematics. Though a participant in the logical empiricist movement during the Vienna years, Gödel thought that Carnap's approach to mathematics could be refuted. The alleged proof (Gödel 1995) was not published in Gödel's lifetime and remains controversial.

Kurt Grelling (1886–1942)

Grelling was born in Berlin and took his doctorate in Göttingen under Hilbert. With Leonard Nelson he developed a famous semantic paradox that bears their names. He was one of the founders of the Berlin Society for Empirical (later Scientific) Philosophy. Grelling died in the Holocaust because for bureaucratic and political reasons news of an academic appointment in the U.S. reached him too late.

Adolf Grünbaum (1923–)

Grünbaum moved from his native Germany as a teenager, studied under Hempel at Yale, and spent the bulk of his career at the University of Pittsburgh, where he founded the Center for Philosophy of Science. The major themes of his work have been philosophy of space and time, rationality, and psychoanalysis.

Hans Hahn (1879–1934)

Hahn, a distinguished mathematician, took his doctorate in his native Vienna in 1902 and began teaching there in 1905. He was part of a group with Frank, Neurath and others that discussed logical and methodological issues prior to World War I. After teaching at Czernowitz (now in Ukraine) and Bonn he was given a chair in mathematics at Vienna in 1921. He was instrumental in bringing Schlick there in 1922 and so was called by Frank "the actual founder of the Vienna Circle" (Stadler 1997/2001, 642). His most famous student was Gödel.

Olaf Helmer (1910–2011)

Helmer took a doctorate in his native Berlin under Reichenbach and a second doctorate under Susan Stebbing in London. He collaborated with other logically minded philosophers. Indeed, the team of Hempel, Helmer, and Oppenheim became known as "H_2O". The bulk of his career was spent at the Rand Corporation.

Carl G. Hempel (1905–1997)

Born just north of Berlin, Hempel studied at both Göttingen and Berlin. Most of his doctoral work was completed under Reichenbach when the latter was forced to leave Germany. Hempel taught at a number of American universities, most famously at Princeton and the University of Pittsburgh. He was the doctor father of many prominent philosophers of science, and his work focused on confirmation, explanation, and concept formation.

Richard Jeffrey (1926–2002)

This American logician and philosopher of science earned an MA with Carnap (with whom he later collaborated) and a PhD with Hempel (with whom he was for many years a colleague and close friend at Princeton). He developed Jeffrey conditionalization (see below) and defended probabilism.

Kurt Lewin (1890–1947)

Born in what is now Poland, Lewin took his doctorate in Berlin in 1916. He lectured there in both philosophy and psychology until 1933 when he emigrated to the U.S. via England. Thereafter he taught at a number of American universities including Cornell, Iowa, MIT, and Duke. Credited with founding modern social psychology, he laid the foundations for what is now called sensitivity training as a way to combat religious and racial prejudices.

Richard von Mises (1883–1953)

Born in what is now Ukraine, Richard von Mises is the brother of the economic and political theorist Ludwig von Mises. Richard was a

polymath who ranged over fields as diverse as mathematics, aerodynamics, philosophy, and Rilke's poetry. He finished his doctorate in Vienna. He was simultaneously active in Berlin, where he was one of the developers of the frequency theory of probability along with Reichenbach, and in Vienna, where he participated in various discussion groups that constituted the Vienna Circle. Eventually it was necessary to escape, first to Turkey, and eventually to MIT and Harvard.

Charles W. Morris (1901–1979)

Morris was an American pragmatist and philosopher of language at the University of Chicago when Carnap arrived there. These two, together with Neurath until the latter's death, were the chief editors of the *Encyclopedia of Unified Science*. After Carnap left Chicago, Morris moved to the University of Florida.

Otto Neurath (1882–1945)

This Austrian philosopher of science and sociologist took his doctorate in political science in Berlin. A member of the First Vienna Circle and a leader of the "left" wing of the Vienna Circle, he was also politically active. He was a significant museum director, and as part of this developed the ISOTYPE picture language. His main philosophic themes were physicalism, anti-metaphysics, and the unity of science. He was the Editor-in-Chief of the *Encyclopedia of Unified Science* until his death. Eventually he fled to the Netherlands and from there to England.

Paul Oppenheim (1885–1977)

A successful industrialist and heir to a substantial fortune, Oppenheim was trained in his native Germany in chemistry and philosophy. He was a close friend of Einstein, and helped to initiate the Berlin Society for Empirical Philosophy. Oppenheim collaborated with many important logicians and philosophers of science both in Europe and the U.S. He also helped many to escape Nazi oppression,

and continued to help in a variety of ways even after he settled in Princeton in 1939.

Karl Popper (1902–1994)

Born in Vienna and with a doctorate there, Popper was intensely engaged in discussions with members of the Vienna Circle. His main philosophical work, *The Logic of Scientific Discovery* (1935/1959), appeared in a series edited by Schlick and P. Frank. He did not however, regularly attend meetings of the Vienna Circle and generally considered himself an outsider. Later he claimed to have "killed" logical positivism. From Austria Popper escaped to New Zealand and eventually to the London School of Economics, where he was knighted for his political writings.

Hilary Putnam (1926–2016)

This American philosopher of science, mathematics, mind and language earned his doctorate under Reichenbach at UCLA and subsequently taught at Princeton, MIT, and Harvard. He was originally a metaphysical realist, but then argued forcefully against it. He has continued the pragmatist tradition and been politically active, especially in the 1960s and 70s.

W.V.O. Quine (1908–2000)

Born in the U.S., Quine took his doctorate and spent his entire career at Harvard. In 1932–33 he visited the Vienna Circle and then Carnap and Warsaw. For the next six years, he said, he was a disciple of Carnap's and even after they began to disagree, Carnap set the agenda. Eventually they clashed over analyticity, modality, and intensional contexts generally. Many similarities of view with Neurath are apparent, especially on the issues of holism, underdetermination, and naturalism in epistemology.

Hans Reichenbach (1891–1953)

Reichenbach was born in Hamburg and, after immersing himself in mathematics, physics, and philosophy, took his doctorate in Erlangen,

Germany. He was a founder and the leader for the Berlin Society for Empirical (later Scientific) Philosophy. In 1933 he was forced to leave Berlin. He went to Turkey and then in 1938 to UCLA. Among his many students were Hempel, Putnam, and W. Salmon, and so almost all philosophy of science in the U.S. can trace its academic lineage to Reichenbach. Though interested in social and educational reform, he worked primarily in philosophy of physics. He developed and defended a frequency theory of probability, and emphasized both scientific realism and the importance of causality and causal laws.

Wesley Salmon (1925–2001)

Salmon was born in Detroit and, after an initial interest in theology, earned his PhD under Reichenbach at UCLA. He taught at a number of universities including Brown, Indiana, Arizona, and Pittsburgh. His interests centered on causality and explanation, and his statistical relevance model of explanation can be thought of as addressing and in large measure resolving the problem of the single case in frequency theories of probability.

Moritz Schlick (1882–1936)

Schlick was born in Berlin and eventually took his doctorate there in mathematical physics under Max Planck. He taught at a number of German universities before he was, at the instigation of Hans Hahn, called to the Chair in the Philosophy of the Inductive Sciences at Vienna, a chair that was previously held by Boltzmann and Mach. Schlick was one of the first philosophers to write about Einstein's relativity theory. He was close to Wittgenstein and one of the conduits for the latter's strict verificationism. His work ranges from space and time to general epistemology and ethics. In 1936 he was assassinated on the steps of the university by a deranged student.

Wilfrid Sellars (1912–1989)

Wilfrid Sellars was the son of well-known philosopher, Roy Wood Sellars. Wilfrid studied at Buffalo, Oxford, and Harvard before

teaching at Iowa, Minnesota, Yale, and Pittsburgh. He was a close associate and collaborator with Feigl at Minnesota. (He once said that he and Feigl were for years discrete parts of a single entity.) He defended scientific realism, pragmatism, and naturalism, and his philosophy of language drew heavily on Carnap's *Logical Syntax* (1934/1937)

Alfred Tarski (1901–1983)

Born and educated in Warsaw, Tarski earned his doctorate under Lesniewski. He happened to be visiting the U.S. when Poland was invaded and so avoided the fate of so many of his colleagues. He taught at the University of California at Berkeley for more than 30 years. While it is unclear whether he should be counted as a logical empiricist, he visited the Vienna Circle and hosted its members in Warsaw, and his "The Concept of Truth in Formalized Languages" (1936/1956) was very influential on Carnap and on the development of semantics among the logical empiricists generally.

Friedrich Waismann (1896–1959)

Waismann was born in Vienna and earned his doctorate there under the direction of Schlick in 1936. From 1926 to 1933 he held discussions with Wittgenstein, generally in the company of Schlick, but also sometimes Carnap or Feigl. Waismann kept detailed minutes of these conversations. At one point he and Wittgenstein contemplated a joint book, but Wittgenstein later changed his mind. Besides the printed text of the *Tractatus* these conversations were the main conduit of Wittgenstein's ideas into the Vienna Circle. In 1937 Waismann was able to emigrate to England. After a couple of years at Cambridge, where he was shunned by Wittgenstein, he moved to Oxford, where he taught until his death.

Ludwig Wittgenstein (1889–1951)

Born into an immensely wealthy Viennese family, Wittgenstein studied at Cambridge from 1911, where he formed friendships with

Russell, Keynes, and Moore. His *Tractatus Logico-Philosophicus* (1921/1922), which among other things tries to show that logic has no content, was enormously influential on many logical empiricists. Wittgenstein continued to spend much of his time in Austria working variously as an elementary school teacher, a gardener, and as an architect of a house for his sister in Vienna. While there he held influential discussions with Schlick, Waismann, and others. From 1930 he held teaching posts at Cambridge and increasingly distanced himself from the logical empiricists. His later work focused on ordinary language and inspired many other philosophers as well.

4. Issues

It is not possible in an essay of this scope to trace all the issues that the logical empiricists addressed or even to treat any one of them with completeness. What is possible is to highlight some salient issues, clear away some misconceptions about them, and sketch a bit how those issues were developed over time. The first is a related set of concerns: empiricism, verificationism, and anti-metaphysics. The second is the logical empiricists' treatment of logic and mathematics as analytic. Third is the related issues of the unity of science and reduction. And finally, comes the issue of probability. Given what has already been said, the reader should be aware that none of the doctrines discussed below was shared by all members of the logical empiricist movement.

4.1 Empiricism, Verificationism, and Anti-metaphysics

Since antiquity the idea that natural science rests importantly on experience has been non-controversial. The only real questions about the sources of scientific knowledge are: Are there parts of science that do not rest on experience or rest also on something other than experience? If so what account can we give of those parts? And to the extent that science

does rest on experience how can we know that it does? There is another question about science related to these, though not strictly about the sources of science, and that is: Why, in making claims about the world, should we be scientific as opposed to say mystical? The difficulty is that any scientific answer to this last question would reasonably be thought to beg the very question it purports to address.

Long before the twentieth century the prevailing opinion was that Euclidean geometry, standard mathematics, and logic did not rest on experience in any obvious way. They were largely presupposed in our empirical work, and it was difficult to see what if anything might disconfirm them. Geometry was a special case and might be handled in different ways that we shall not discuss here. That leaves logic and mathematics.

If Frege and Russell were right, then mathematics could be thought of as expressing no more than logical truths and handled in whatever way logic was to be treated. For Frege both mathematics and logic were analytic, but that, even if true, does not provide the needed answers. Wittgenstein's no-content theory of logic suggested that all of the real claims, the ones that had genuine content, could be appropriately supported by experience, and the logical and hence mathematical claims had no content to support. This seemed to open the way for a thoroughgoing empiricism in which the logical and mathematical fit in with the ordinary claims of physics and biology in a harmonious way. The next subsection about analyticity discusses the question of whether the needed distinctions can be drawn.

In developing his theory of types Russell said in effect that some expressions that seem to be sentences in fact say nothing at all. This is because, despite appearances, they are not grammatically well formed. Wittgenstein found this suggestive. In the *Tractatus* he suggested that much else was nonsense as well including traditional metaphysics and

supposed claims about the "higher". When in late 1929 Wittgenstein proposed (Waismann 1967/1979), in conversations with Schlick and Waismann, a strict verificationism as a basis for identifying the legitimate parts of discourse, this seemed to the logical empiricists to be a very attractive tool for setting aside the unscientific parts of philosophy.

This does not mean, however, that all logical empiricists or even all members of the Vienna Circle accepted the strict verificationist view that in order to be meaningful a claim must be implied by a finite number of observation sentences. Even though those observation sentences need not be true, this view had the drawback that so-called laws of nature would not be meaningful on this criterion. Schlick was prepared to bite the bullet and hold that laws were not statements at all but principles of inference. Others were not prepared to go so far and sought more liberal formulations. This more liberal or "left" wing of the Vienna Circle included Carnap, Philipp Frank, Hahn, and Neurath. Carnap does not seem to have been a strict verificationist even in the *Aufbau* (1928/1967).

Over the years a great many different formulations of verificationist principles ensued. Most of them came to a bad end rather quickly, and this is sometimes taken as a convincing argument that any form of verificationism is utterly misguided. Perhaps, but we should be cautious. There are undoubtedly many different features joined in any one of the proposals, and even a sequence of failures may not show where to place the blame. The central idea behind verificationism is linking some sort of meaningfulness with (in principle) confirmation, at least for synthetic sentences. The actual formulations embodied not only such a link but various particular accounts of confirmation as well. Now confirmation is a complex matter, and it is unlikely that we shall have the final satisfactory account any time soon. This should not persuade us, however, that there are no satisfactory accounts of confirmation any more than our current lack of the final physics should convince us that there are no physical facts

of the matter. So even a string of failures in formulating verificationist principles may mean no more than that the embedded accounts of confirmation are too simple but the link between meaningfulness and confirmation is nevertheless sound.

Even if we set this caution aside, there may be parts of persistently employed strategy that lead to persistent failure. These parts and failures might be avoidable. To see how this may be so we will compare what is perhaps the most famous formulation of the verificationist principle, in Ayer 1936, with a later one, in Carnap 1956. A.J. Ayer had visited the Vienna Circle from late 1932 on into 1933, returning home for the summer term. While in Vienna he attended meetings of the Circle and overlapped for five weeks with Quine. Neither Carnap nor Neurath were there at the time, so the left wing of the Circle was not fully represented. When Ayer returned to England he published *Language, Truth, and Logic* in 1936. Even immediately it was widely discussed, and after the war sales were spectacular. For many in England this book was the epitome of logical positivism and remains so.

Ayer was careful to restrict his criterion of meaningfulness to synthetic sentences and to demand only in principle confirmation. And the formulation seems very natural: Confirmation is a feature that applies to sentences (or groups of them) and not to sub-sentential parts, and for an empiricist the content that a synthetic sentence has would be empirical content. So it would seem that to have empirical content a sentence, A, should either directly imply some observational sentence or add to the observational content of some other sentence, B. That is, the conjunction of A and B should imply some observational sentence not implied by B alone. This formulation may be natural, but it is also fatally flawed. It would declare any sentence whatsoever as meaningful: For any sentence A and any observation sentence O, A would be meaningful because it could

be conjoined to $A \supset O$. The latter would not in general imply O, but the conjunction would.

Other more elaborate formulations followed along the same lines, and other more elaborate counterexamples appeared just as fast. Hempel reviewed the situation twice within about a year (Hempel 1950 and 1951). First he concluded that it was a lively and promising line of research and later concluded that it was not promising at all. In retrospect it may be that the problems arise because we were led by the fact that confirmation is a feature that applies to whole sentences into thinking that the level at which to apply the criterion was the level of whole sentences. Now a sentence with meaningless parts might well pass some test especially if the test involves its being combined with other sentences that can have meaningless parts. So one way to avoid this difficulty is to try to find a formulation that applies the test at the level of basic expressions, those that can be thought of as "not having parts" so to speak.

This is the strategy that Carnap employed in "The Methodological Character of Theoretical Concepts" (1956). Observational terms are assumed to have empirical content. Logical terms are assumed to have none. And all defined terms are assumed to be replaced by their definitions. If for some basic, non-logical term there is a sentence that contains that term as its only non-logical element and if that sentence implies some observation sentence, then that sentence has empirical content and so does its only non-logical term. If we have established that each term from some set, K, is empirically significant we might test still further terms by seeing whether those further terms can add to what is sayable with terms from K. Carnap's actual definition is quite complicated, but it does seem to avoid the difficulties of its predecessors. It also allows an account of why those predecessors ran into trouble, viz., that they applied at the level of whole sentences (naturally enough) rather than to elementary terms.

Not long after Carnap's definition was published David Kaplan devised what seemed to be counterexamples. They became fairly well known, but they were not published until 1975. Shortly thereafter it was shown (Creath 1976) that either Carnap's definition is not open to the counterexamples as presented or it can be patched in a very natural way so that it avoids them. This does not show that there are no counterexamples or that there are no other features of the definition to which one might object. But it does show that the situation is not as dire as Hempel supposed in 1951.

We need to address another issue in considering verificationism, the persistent criticism that it is self-undercutting. The argument for this claim goes like this: The principle claims that every meaningful sentence is either analytic or verifiable. Well, the principle itself is surely not analytic; we understand the meanings of the words in it perfectly well because we understand our own language. And we still do not think it true, so it cannot be true in virtue of meaning. And it is not verifiable either (whatever we choose 'verifiable' to mean).

This sounds more compelling than it is. Ayer understood the principle to be a definition, defining a technical term, 'meaning'. If so, then the sentence expressing the principle would indeed be analytic. So the self-undercutting charge strictly fails. But so construed and with nothing else said about it the principle would not have the same punch as before. Why should a metaphysician care whether his or her utterances lack some technical feature?

Carnap explicitly takes up the "self-undercutting" charge against verifiability in *Philosophy and Logical Syntax* (1935), and he is not interested in introducing a new technical term, 'meaning', so or in denying this new technical property to unverifiable sentences. Carnap is careful to distinguish the language for which the verifiability principle is given from

the meta-language in which we talk about that language. This meta-language would be the language in which the principle would be expressed. This may seem to offer another strategy against the "self-undercutting" charge because the principle applies to a different language than that in which it is expressed. This is not Carnap's strategy. Carnap fully understands that if the general verificationist strategy is followed, there will also be a verificationist principle expressed in the meta-meta-language governing the meta-language.

Carnap's real defense of the principle was achieved by changing the nature of the discussion. By 1935 Carnap had introduced an important new element into his philosophy called the Principle of Tolerance. Tolerance is a radical idea. There is no uniquely correct logic (1934/1937 xiv–xv). Empiricism is a convention (Carnap, 1936/1937 33). Perhaps more precisely each of the various versions of empiricism (including some sort of verificationism) is best understood as a proposal for structuring the language of science. Before tolerance, both empiricism and verificationism are announced as if they are simply correct. Correspondingly, what Carnap called metaphysics is then treated as though it is, as a matter of brute fact, unintelligible. But what is announced thus dogmatically can be rejected equally dogmatically. Once tolerance is in place, alternative philosophic positions, including metaphysical ones, are construed as alternative proposals for structuring the language of science.

None of them is the uniquely correct one, and no theoretical argument or evidence can show that it is. Nor can theoretical arguments or evidence show that it is false. Neither proposals nor languages are the sort of thing to be true or false. Instead, proposals call for practical decisions and practical arguments rather than for theoretical reasons or evidence. Carnap believes that there are indeed very good practical reasons for adopting the proposal of verificationism, for choosing a language of science in which all substantive (synthetic) claims can, at least in principle, be brought

before the court of public experience. The reason is that if we do not require this, the result is "wearisome controversies" that there is no hope of resolving. That, he thinks, is the sad history of attempts to get beyond science, and it is just too painful.

If the proposals constituting some version of verificationism are adopted, then in the language thus constituted it will be analytically true that there are no synthetic sentences that are both unverifiable and meaningful. The notion of meaning here is not some new technical invention. Rather, 'meaning' is used in something like the ordinary sense. No grammatically well-formed sentence of this new language violates the verifiability principle. And the principle itself is completely safe.

Thought of in this way the verifiability principle does not describe natural language, it is not intended to. It is intended to reform language to make it a more useful tool for the purposes of science. Carnap is under no illusion that natural languages are free from metaphysics. Nor is he under the illusion that defenders of the sort of metaphysics he targets will readily step up to the challenge of presenting precise rules of grammar and inference.

There is one other change that tolerance brings to Carnap's own vocabulary. Before tolerance, verificationism is stated in such a way that violations would count only as unintelligible gibberish. With tolerance in place, Carnap is prepared to imagine non-empiricist languages, though of course he thinks they are very unwise. So instead of saying that sentences in non-empiricist languages are meaningless, he says that they are empirically meaningless. And that has a very different flavor. There is no weakening of his defense of empiricism, but it is put on a somewhat different footing.

4.2 Analyticity

Logic, mathematics, and mathematical geometry had traditionally seemed to be confirmationally "different". Indeed it is hard to indicate any conditions under which any parts of them would be disconfirmed. Leibniz had called them truths of reason. Hume said that they represented relations of ideas. Kant had held that the truths in these areas were a priori. Mathematics and geometry were not analytic for Kant, but logic was. Kant had two criteria of analyticity, apparently thinking them equivalent. First, in subject-predicate sentences, an analytic sentence is one in which the concept of the predicate is contained in that of the subject. Second, an analytic sentence is one whose denial is self-contradictory. This seems to include not only the sentences whose surface logical form would be of the required sort but also those that can be got from such logical truths by making substitutions that were conceptually equivalent. The more modern rough analog of this is to say that the analytic sentences are those that are true in virtue of logic and definition.

Frege certainly developed logic beyond that which was available to Kant, but he did not think of himself as changing the analytic status of it. Logic is after all the only avenue we have for giving meaning to the notion of (logical) contradiction. Of course Frege also attempted to reduce mathematics to logic (including both first and second order logic), and insofar as that reduction was successful it would have implied that mathematics was analytic as well. Frege said little of geometry, but for him it was synthetic a priori.

Carnap had not only studied with Frege, but like many of the logical empiricists he had started out as a neo-Kantian as well. So especially in view of Russell's relatively more successful attempt at reducing mathematics to logic, it was perhaps natural that Carnap would consider both mathematics and logic as analytic. Geometry could be handled in several different ways that we will not discuss here. But from fairly early on there was widespread agreement among the logical empiricists that

there was no synthetic a priori, and that logic and mathematics and perhaps much else that seemed impervious to empirical disconfirmation should be thought of as analytic. The point of drawing the analytic-synthetic distinction, then, is not to divide the body of scientific truths or to divide philosophy from science, but to show how to integrate them into a natural scientific whole. Along the way the distinction clarifies which inferences are to be taken as legitimate and which are not. If, as Carnap and Neurath were, you are impressed by Duhemian arguments to the effect that generally claims must be combined in order to test them, the analytic-synthetic distinction allows you to clarify which combinations of claims are testable.

If analytic, a sentence is true in virtue of the conventions of language. In saying that, however, we must pause to confront two widespread confusions. First, Quine alleges (1963, 385f) that the notion of analyticity was developed and purports to explain for both Kant and Carnap how certainty is possible. In fact certainty has little or nothing to do with analyticity for the leading logical empiricists. In saying that such claims are based on convention they were explicitly calling attention to the revisability of conventions and the sentences that owed their meanings to those conventions. Second, nowadays any talk of convention is likely to prompt the response: "But that cannot be! No proposition can be made true by our conventions or decisions." Unless it is a proposition about conventions, this second sentence of the response is true. But it is also completely irrelevant. Analyticity applies to *sentences* rather than propositions. Our conventions and decisions can and do affect what expressions mean and thus what sentences mean. Once the meaning is specified, it may well be that any sentence that has this meaning would be true even if, for example, the point masses of the universe were arranged quite otherwise than they in fact are. These are the analytic sentences. No claim is being made that meaning causes anything or that convention *makes* anything true. The "making" image here is out of place. It is just

that in these cases the truth value of the sentence may well be functionally dependent on meaning alone. If it is, then in this special sense, truth value depends on meaning, and that depends on convention. Other sentences whose meanings are specified might well be true or false depending on how things in the external world, so to speak, are arranged. In this other category of sentence the truth value is not functionally dependent on meaning alone. They are the synthetic sentences. Now this puts matters extremely informally. But at least the nature of the confusions over certainty and convention should be clear.

In the *Logical Syntax of Language* (1934/1937) Carnap defined 'analytic' in a new way in order to circumvent Gödel's incompleteness results. The method used was to distinguish between a derivation relation (the relation that holds between some premises and what can be got from them in a finite number of steps) and a consequence relation. The latter is an essentially semantic relation that holds between some premises and some other claim such that on all valuations under which the premises are all true, so is that other claim. This definition bears a stronger resemblance to Tarski's account in (Tarski 1936b/1956). In any case, Carnap is able to show that for any sentence of pure mathematics either it or its negation is a consequence of the null set of premises. This leaves Gödel's results completely intact as they concerned what is provable, that is, derivable from the null set of premises or from any one consistent axiomatization of mathematical truths.

As noted above, another innovation of *Logical Syntax* is the Principle of Tolerance. While it reflects a long-standing attitude on Carnap's part, the principle itself is new. Later Carnap was to say that the Principle of Tolerance was "perhaps better called the principle of conventionality" (Carnap 1942, 247), that is, the conventionality of linguistic forms. Tolerance stabilizes the verification principle as well as Carnap's

empiricism, and it reinforces the idea that the analytic-synthetic distinction is always relative to a particular language (Creath 2009).

In the late 1950s Carnap began exploring (1963a and 1966) how a notion of analyticity might be developed for novel theoretical terms where the theories in which those terms are embedded are presented by means of a system of postulates. It is not clear that the account he developed was intended to supersede his earlier account. In any case Carnap's suggestion is as follows (where for convenience terms are used autonymously): Let T be the totality of theoretical postulates, and C be the totality of mixed sentences (the sentences of the theory containing both antecedent and novel terms). Also let $R(TC)$ be the Ramsey sentence for TC, that is, the result of replacing each of the non-observational terms in TC with predicate variables and closing that open sentence with corresponding existential quantifiers. $R(TC) \supset TC$ can, Carnap says, be thought of as the analytic sentence for the theory, that is, a sentence that gives to the theoretical terms of TC their meaning. Over the last decade, this idea of Carnap's has provoked considerable discussion that has not yet been resolved. Whatever worries there may be concerning this part of Carnap's view, they are distinct from the more famous concerns raised by Quine.

Quine began having doubts about analyticity about 1940, though he seems not to have been firmly committed against it until later. In any case his doubts were not published until 1951 in his famous paper "Two Dogmas of Empiricism". Quine's readers have understood his arguments in many different ways. The most general form of his complaint is that 'analytic' so far lacks the appropriate tie to observational criteria that Carnap's own account of theoretical terms in empirical science would demand. More specifically, where there has been an attempt at such a general criterion it has resulted in either a "drastic failure as tended to admit all or no sentences as analytic, or there has been a circularity" (Quine 1963, 404) of a kind that defines 'analytic' in terms that themselves lack the appropriate

empirical criteria and so can be accounted for only by appeal to analyticity itself.

This complaint falls far short, as Quine well understood, of a proof that Carnap's appeal to analyticity was doomed. First, it relies on the demand that theoretical terms must satisfy some empirical significance criterion. Many people at the time, including some who followed Quine in rejecting analyticity, also rejected any general empirical significance demand for theoretical terms. Second, one could accept the demand for theoretical terms in physics or chemistry and deny, as Carnap did, that the demand applied to his own work. This is because Carnap saw himself as working in an area within metamathematics rather than in empirical linguistics. Third, Quine did not pretend to have considered all of the possibilities for the explication of analyticity. And so it may be possible to meet Quine's demands to the extent that they are legitimate. Fourth and finally, Quine seems in *Roots of Reference* (1974) to have provided an explication for 'analytic' that meets his demand for empirical/behavioral criteria without inducing either the drastic failure or the circularity envisioned above.

There is another somewhat independent thrust to Quine's campaign against analyticity. In the last section of "Two Dogmas" (1951) Quine gives an extremely attractive sketch for an alternative epistemology that apparently makes no appeal to analyticity. Insofar as that sketch can be filled out successfully it would constitute a dispensability argument against analyticity. Whether it can be thus filled out, however, remains to be seen.

Quine's other provocative theses, including especially his claims about the indeterminacy of translation, while relevant to his assessment of analyticity, would carry us too far afield to consider their ramifications here. As with most topics in philosophy there is no uniform agreement in the literature as to whether the notion of analyticity is or can be made

sufficiently clear for use in scientific philosophy. Nor is there such agreement that Quine's epistemological sketch can be satisfactorily filled out. Both approaches have their defenders and their detractors. But between them they seem to be the most promising avenues for integrating the logic-mathematical part of science with the more straightforwardly empirical parts. Since Carnap is and Quine can be argued to be within the logical empiricist tradition, this progress toward such unification can be counted as part of the legacy of the movement.

4.3 Unity of Science and Reduction

The commitment of some of the logical empiricists to the unity of science has been in recent years often discussed but less often understood. One hears in conversation that it was a sort of rearguard action designed to preserve as much as possible of a phenomenalist version of ontological reduction. One reads in print that it can be refuted by the obvious fact that the various sciences have quite distinct theoretical vocabularies (Suppes 1978). Both reactions are misplaced.

It was the left wing of the Vienna Circle, and above all Otto Neurath, that championed the unity of science. They also promoted physicalism, anti-foundationalism, and a generally naturalistic viewpoint. A main focus of their activities from the late 30s was *The Encyclopedia of Unified Science* edited by Neurath in Europe and Carnap and Charles Morris in Chicago. A great many philosophers of many different persuasions participated in that project. The project may have been unified science, but they did not have a completely unified view of what that project was. Here we will discuss the Neurath and Carnap versions of it to see what their central concerns were.

Neurath seems to have had two primary motivations to advance under the banner of the unity of science. First, he was concerned that there be no a priori methodological cleavage between the natural and the social

sciences. On the social scientific side he was concerned that these sciences not condone some private, mysterious mode of insight (empathy) whose results could not be checked against more ordinary public observation. Such a methodology would be a harbor for metaphysics. On the natural scientific side, he was concerned to point out that, for Duhemian and other reasons, the situation is much messier than is sometimes supposed, and so invidious comparisons by natural scientists at the expense of social science were unwarranted.

Second, because Neurath was socially and politically engaged he was concerned that the various sciences be connected in such a way that they could be used *together* to solve complex human and social problems. For this, considerable overlap of vocabulary was needed, and this he called a "universal jargon".

In recent years it is sometimes claimed that Neurath meant by the unity of science what some contemporary philosophers have defended as the disunity of science. One cannot rule this claim out a priori. But the often substantial differences among the current defenses of disunity make evaluating this claim difficult. It is fair to say, however, that Neurath was suspicious of grand hypotheses, familiar since the 19th century to derive all of chemistry, biology, psychology, and the social sciences (in that order) from a few basic principles of physics. It is unclear whether this stems from a general opposition to system building, since he was eager to develop inferential connections among the various sciences. Perhaps this is better expressed as an opposition to *speculative* system building and to the idea that there is only one way of systematizing our science than to systematicity as such.

Carnap's position on unity is different from Neurath's, but they overlap. Carnap distinguished the unity of the language of science from the unity of the laws of science. He wanted to defend the former and to say what

would be required for the latter. As far as the unity of the language of science, Carnap did in the *Aufbau* try to initiate a program for defining all of scientific concepts on the basis of a very small number of basic concepts, perhaps only one basic concept. That does afford a certain conceptual economy, but it is now generally held by Carnap scholars (see especially Friedman 1987 and Richardson 1998) that ontological reduction and reduction to a phenomenalist basis was far from his motive. Carnap explicitly acknowledged that another system of definitions, one with a physicalist basis, might also be possible. Instead of ontological economy and a phenomenal basis, Carnap's project seems to have been the more Kantian one of indicating how semantic intersubjectivity is possible: How can it be that, even though I have only my own experiences and you have only yours, we can nevertheless share a common body of concepts? The answer is given in terms of shared inferential structure and identifying any given concept with a unique place within that shared overall structure. This is a highly holistic conception of concepts and it depends on thinking of the body of scientific commitments as a whole, as a unity.

The *Aufbau* was largely drafted before Carnap joined the Vienna Circle. Once there and under some influence from Neurath, Carnap campaigned more insistently for physicalism and for the unity of science. They seemed often to be two sides of the same coin. From 1933 onward there was a succession of monograph series with 'Unified Science' in the title. Until his death in 1945, Neurath was in each case the main editor and Carnap either the associate editor or one of the associate editors. *The International Encyclopedia of Unified Science*, begun in 1938 is undoubtedly the most famous of these. Carnap's own essay on this topic "Logical Foundations of the Unity of Science" (1938) was printed as part of the very first number in the encyclopedia.

The dates here are relevant because by the time of this essay Carnap had already decided (Carnap 1936–37) that theoretical terms could not in

general be given explicit definitions in the observation language even though the observation reports were already in a physicalist vocabulary. The partially defined theoretical terms could not be eliminated. This seems to have caused Carnap no consternation at all, and it never seems to have occurred to him that there was any conflict whatever between this result and the unity of science. This is because by this point the elimination of concepts was not the point of the exercise; their inferential and evidential integration was.

In the 1936–37 article, "Testability and Meaning" Carnap called the partial definitions themselves "reduction sentences" and the system of definitions of theoretical terms, both partial and complete, as a reduction of the theoretical terms to the observational basis. Plainly he means by the word 'reduction' something other than what we currently mean, not that there is anything univocal about current uses of the word. By 'reduction' of vocabulary A to vocabulary B Carnap means the specification of the inferential relations that would allow us to say what sentences or combinations of sentences in A would count as evidence for sentences in B.

This is also the key to what Carnap means by the unity of the language of science. The language of science is unified, no matter how different and exotic its various technical vocabularies may be, when each of its terms is reduced to (can be tested in) a common public observation vocabulary. The call for the unity of the language of science, then, amounts to no more than the demand that the various claims of the separate sciences should be publically testable in a common observation language. Controversies will of course arise as to what the observational vocabulary should be and what are the acceptable forms of linkage. Carnap's demand for unity in the language of science abstracts from those controversies to concentrate on the goal of public testability. That does not seem to be an unreasonable demand.

The unity of the language of science so far discussed is quite a different issue from the unity of the laws of science. And Carnap's attitudes toward them are quite different. The latter issue concerns the extent to which the laws of one special science can be inferred from those of another. Carnap tries to articulate what would be involved in such a unification, but he nowhere says that such a unity is either possible or mandatory. Finding any sort of inferential connections among sets of laws would be welcome of course. But the question of how much unity there is, if any, among the various sciences is an empirical question that philosophers are ill equipped to answer. Philosophers should not make pronouncements, especially in advance of having putative laws in hand, either that scientific laws are unified or that they are not. A certain modest deference to the empirical facts that philosophers generally do not have, again, does not seem unreasonable.

Taking unity as a working hypothesis, as some philosophers have done, amounts to *looking for* inferential and nomological connections among various sets of laws, but not to the assertion that such connection will be found. Even if we accept the idea that such connections would be welcome if found, the question of whether one should spend significant effort in looking for them is not thereby answered. That would be a difficult and delicate practical question of how to apportion one's research effort that for the purposes of this essay we must set aside.

4.4 Probability

There are two broad approaches to probability represented in logical empiricism. One of these, the so-called frequentist approach, has an extensive 19th century history and was further developed from about 1920 onward by Richard von Mises and Hans Reichenbach. The other is the epistemic approach to probability. This goes back at least to Laplace at the end of the 18th century. In the 20th century Rudolf Carnap, who explored

what he called logical probability, and Frank Ramsey and Richard Jeffrey whose accounts can be distinguished from Carnap's and are often called subjective probability, all defended the epistemic approach. While Ramsey visited the Vienna Circle he was not much influenced by its members on these matters. By contrast, Jeffrey studied and later collaborated with Carnap but also made significant contributions of his own.

It is natural to begin thinking about probabilities with a simple mathematical account that takes as its point of departure various games of chance involving cards, dice, or coins. Bettors have long noted that some outcomes are much more likely than others. In this context it is convenient to take the probability of a kind of outcome to be the ratio of such outcomes to all possible outcomes. Usually for reasons of symmetry in the physical set up, the possible outcomes are assumed to be equally likely. Where that assumption happens to be true or nearly so the empirical results of, say, a great many throws of a pair of dice tends to be close to what the simple mathematical account would suggest. Conversely, where the outcomes deviate from the expected ratios, bettors begin to suspect that the dice, coins, and cards (or the manipulations of them) are not all that they seem. The suspicion is that the outcomes are not equally likely and that the simple mathematical account does not apply.

These facts suggest both two limitations of the simple account and the beginnings of a way around them. The first limitation is that the account applies only where the outcomes can be partitioned into alternatives that are equally likely. This is not the case when dice are loaded or in such real world cases as radioactive decay or weather forecasting. A second limitation is that the account, in describing the possible outcomes as equally likely, implicitly appeals to the very probability notion for which clarification was sought. The realization that we can sometimes discover the falsehood of the assumption of equal likelihood and make a much more reasonable estimate of probability by making a large number of trials

is very suggestive. And from his dissertation onward Reichenbach worked out a variety of imagined physical models that could guide ones thinking about probability in useful ways. The result is what is often called the frequency theory of probability (or sometimes the statistical frequency theory or the limit frequency theory).

Even a perfectly fair coin in an odd number of flips will never result in exactly the same number of heads and of tails. When the coin is fair and the number of flips is even, an outcome perfectly balanced between heads and tails is not guaranteed either. So, even on the assumption that the probability of the coin's coming up heads does not change over the course of the trials, we need to be cautious. A larger number of flips might make us more confident that the ratio we have seen is close to the "actual" value, but there is no finite number of flips after which we can say that the observed ratio is exactly right. We will never make an infinite number of flips either, and in actual cases a large finite number of flips might so erode the coin as to bias the coin and discredit the result. Notwithstanding these limitations on an actual series of trials one can imagine an infinite series of trials and define a notion of probability with respect to it. This raises its own difficulty, namely that ratios are not defined for infinite collections. They would be defined, however, for any finite initial segment of such an infinite series, thus giving a sequence of ratios. If this sequence of ratios settles down on a limit, the probability of the coin showing a head given that it has been flipped can be defined as the limit of the ratio of heads to total flips as the number of flips goes to infinity.

While probability thus defined has a somewhat counterfactual character, that is not an obvious defect. Moreover, this notion of probability applies perfectly well to biased coins and loaded dice, as well as to radioactive decay. On the surface at least it also seem to avoid using the notion of probability in its own definition, and in these respects it seems to be an important improvement over the simple mathematical model with which

we began. The definition locates the probability objectively "out in nature" so to speak, and this comports well with Reichenbach's scientific realism.

A problem that remained troublesome concerns the fact that one often wants to assign probabilities to particular events, events that in the nature of things cannot be repeated in all their particularity. Thus it is unclear how a frequency theory of probability is to be applied to such individual cases. This is often called the problem of the single case. It is a little difficult to assess how serious this is, because in actual practice we often have no difficulty in making probability assignments to single cases. Suppose we are interested in the probability of rain tomorrow. Tomorrow will never be repeated, and we want to estimate the probability now. What we do is to look back through the records to find days relevantly like today and determine in what fraction of those cases those days were followed by rainy days and use that as our estimate. Even if we are comfortable with this practice, however, it is another matter to say why this should give us a reasonable estimate of the value of the limit involved in a logically impossible infinite sequence. This problem of the single case was much discussed, and Wesley Salmon made progress in dealing with it. Indeed, Salmon's account of statistical explanation can be viewed as a substantial mitigation of the problem of the single case (W. Salmon 1970).

There are residual difficulties in making estimates of the probabilities on the basis of finite evidence. The problem is that even when we are assured that the sequence of ratios has a limit, we have no a priori grounds for saying how close the current ratio is to that limit. We can boldly estimate the limit by means of the so-called "straight rule". This just takes the most recent ratio as the desired estimate. This is a good practical solution where the number of trials is already high, but this does not really say why the estimate should be good, how good it is supposed to be, or how many trials would be high enough. In addition, the straight rule can yield counterintuitive results where the number of trials is small.

Though there are these issues outstanding, frequency theories define a concept of probability indispensable for quantum theories and for a wide variety of other applications in the natural and social sciences. It was not the only concept of probability to be developed by the logical empiricist tradition. The primary other such concept was the epistemic conception of probability. We will begin with Carnap and then move to those who developed a subjectivist account.

Carnap is addressing a different issue than was addressed by von Mises and Reichenbach. Instead of focusing on physical phenomena and ratios within them, Carnap focuses on arguments and takes as his point of departure the widespread conviction that some arguments are stronger, in varying degrees, than others, even for the same conclusion. Similarly some bodies of evidence can give us more reason to believe a given conclusion than would another body of evidence. Carnap sets as his task the development of a quantitative concept of probability that will clarify and explicate these widespread convictions. Such a quantitative concept would be an extraordinarily useful tool, and it would be a useful successor to our ordinary, somewhat scattered notions of confirmation and induction.

Carnap approaches the problem by first considering extremely limited artificial languages and trying to find a confirmation function that will work for that. If he succeeds he would then try to develop an account that would work for a broader and richer range of languages. In this his approach is like that of a physicist developing a physical theory for the highly artificial situation of a billiard table or air track and then broadening the theory to deal with a wider range of cases. In Carnap's case, however, it is somewhat unclear what success would be in an artificial language very much unlike our own. In any case, Carnap is not trying to describe our linguistic habits but to clarify or even to replace them with something more useful.

As early as *Logical Syntax* (Carnap 1934/1937, 244/316–17) Carnap had suggested that Wittgenstein's remarks in the *Tractatus* about ranges (*Tractatus*, 4.463) might be a starting point for thinking about probability. By 1945 Carnap also distinguished the two approaches described here, insisting that they were not competitors but were attempting to explicate two different concepts of probability. One need not choose one as the only concept; both concepts were useful. Reichenbach, by contrast, never conceded that both concepts were needed and insisted that his frequency notion could serve all epistemic purposes for which any notion of probability is needed.

Carnap's general strategy was first to identify a broad class of confirmation functions, as subjectivists Ramsay and de Finetti were also to do, and then find a natural way of limiting this class still further. The confirmation functions have to meet some basic mathematical conditions. The axioms that state these conditions partially define a function, and this function can be interpreted in a number of ways. Carnap himself lists three in Carnap 1950. In (1955), John Kemeny (one of Carnap's collaborators and later a co-inventor of BASIC programming language and still later president of Dartmouth College) gave an argument that persuaded Carnap that it was more fruitful to think of the function as indicating fair betting quotients rather than evidential support. This took Carnap even closer in conception to the work of such subjectivists as Ramsey and de Finetti. Indeed, the discussion of fair betting quotients, and related issues of Dutch book arguments had been initiated by de Finetti.

In *Logical Foundations of Probability* (1950) Carnap had discussed Bayes' theorem and promised to expand the discussion in a second volume. Carnap's interest in Baysianism grew, but that second volume never materialized, quite possibly because rapid development of the field was still under way at the time of Carnap's death. As his work proceeded Carnap tended to explain probabilities by reference to events and

propositions rather than speak overtly about sentences. A similar change appears in the rest of Carnap's work as well. It is not clear, however, whether this amounts to a major change of view or a change in what he sees as the most felicitous mode of expression. As the years progressed Carnap tended to see the remaining differences between himself and his subjectivist co-workers as chiefly differences in emphasis. In any case the subjectivist tradition is now dominant in philosophical discussions of probability (Zabell 2007, 293). Richard Jeffrey whose own work arose out of logical empiricism carried on that tradition for 35 years after Carnap's death. Jeffrey himself made major contributions including a principle for updating ones beliefs when the evidence one learns is not certain. The world knows this principle as "Jeffrey conditionalization"; he called it simply "probability kinematics".

Popper's view of probability, his propensity theory, differs from either of the two approaches discussed above. Unlike the epistemic approach of Carnap and others, Popper was not trying to clarify inductive relations because he did not believe that there are inductive inferences. Theories can be corroborated by their passing severe tests, but they are not thereby inductively confirmed or made more probable. For a discussion of whether there are any significant similarities between Popper's idea of corroboration and the ideas of inductive confirmation that he rejects, see (Salmon 1967, 1968).

Propensities are thought of as tendencies of a physical event or state to produce another event or state. Because propensities are to be features of external events and not, to use Hume's phrase, relations of ideas, the propensity theory and the statistical-frequency theory are sometimes grouped together as accounts of chance. Popper has specifically applied propensities to single non-repeatable events (1957), and that suggests that the concept of propensity does not involve any essential reference to long sequences of events. Popper has also taken propensities as producing

outcomes with a certain limit frequency (1959). This does suggest a rather closer tie to the statistical frequency approach. Later philosophers developed both sorts of propensity theories, single-case theories and long-run theories. (Gillies 2000) And like other approaches to probability and induction all these views remain controversial. While we will not discuss the relative merits of the various approaches further, those who are interested in Popper's views in this area should look at the many papers on probability, induction, confirmation, and corroboration, and Popper's replies, in *The Philosophy of Karl Popper* (Schilpp 1974).

5. Impact

In 1967 John Passmore reported that: "Logical positivism, then, is dead, or as dead as a philosophical movement ever becomes." (1967, 57) Earlier in the same article he had equated logical positivism with logical empiricism, so presumably that was dead too. At that time few would have disagreed with Passmore, even though Carnap was still alive and active. But in speaking of this movement Passmore was referring not to a movement but to specific doctrines, and his interpretation of them was much influenced by Ayer. Even so, Passmore conceded that the movement had left a legacy and that "the spirit which inspired the Vienna circle" persisted. It still does.

Part of the movement's legacy lies in contemporary philosophy of science. In the US nearly all philosophers of science can trace their academic lineages to Reichenbach. Most were either his students or students of his students and so on. His scientific realism inspired a generation of philosophers, even those clearly outside the movement. Even the reaction against various forms of realism that have appeared in recent decades have roots in the logical empiricist movement. Moreover, philosophers of science are expected to know a great deal of the science about which they philosophize and to be cautious in telling practicing scientists what

concepts they may or may not use. In these respects and others contemporary philosophers promote a kind of naturalism, and by so doing they follow both the precept and the example of the logical empiricists.

There are other issues where the legacy of logical empiricism is still visible. Two different approaches to probability are still under discussion. One of them explores the objective chances of external events; this investigation follows in the tradition or the frequency theory of Reichenbach and von Mises. The second approach has an epistemic conception of probability as exemplified by Carnap. S.L. Zabell summarizes the current situation as follows:

> But although the technical contributions of Carnap and his school remain of considerable interest today, Carnap's most lasting influence was more subtle but also more important: he largely shaped the way current philosophy views the nature and role of probability, in particular its widespread acceptance of the Bayesian paradigm (as, for example, in Earman, 1992; Howson and Urbach, 1993; and Jeffrey, 2004). (Zabell 2007, 294)

There is also a continuing concern for how the various sciences fit together. Some have scouted theoretical unification and others a more pluralistic model, just as the logical empiricists did. There was for a while a vogue for the disunity of science. Some even said that their conception of the disunity of science is just what Neurath meant by the unity of science. Parts of the discussion were intended as challenges to logical empiricism, but often the arguments used were pioneered by the logical empiricists themselves.

For the 30 years after Passmore's report metaphysics became ever more visible in philosophy. It was a diverse development, but in the self-conceptions of many of its most prominent practitioners there was no

attempt to shun science or logic or to think that metaphysics had access to facts that were deeper than or beyond those that a proper science could reach. So the metaphysics that blossomed was not necessarily of the sort that Carnap and others combated. Most recently there are some in meta-ontology that want to reconsider and reconnect with Carnap's ontological caution.

Even in its heyday many philosophers who on either doctrinal or sociological grounds can be grouped with the logical empiricists did not see themselves that way. We should not expect philosophers today to identify with the movement either. Each generation finds its place by emphasizing its differences from what has gone before. But the spirit of the movement still has its adherents. There are many who value clarity and who want to understand the methodology of science, its structure, and its prospects. There are many who want to find a natural home within a broad conception of science for conceptual innovation, for logic and mathematics, and for their own study of methodology. And importantly there are those who see in science a prospect for intellectual and social reform and who see in their own study of science some hope for freeing us all from the merely habitual ways of thinking "by which we are now possessed" (Kuhn 1962, 1). These are the motives that define the movement called logical empiricism. As Twain might have said, the reports of its death are greatly exaggerated.

Bibliography

Cited Literature

Ayer, A.J., 1936, *Language Truth, and Logic*, London: Gollancz.
Carnap, R., 1928/1967, *Der logische Aufbau der Welt*, translated by R.A. George as *The Logical Structure of the World*, Berkeley: University of California Press.

Existentialism

First published Mon Aug 23, 2004; substantive revision Mon Mar 9, 2015

Like "rationalism" and "empiricism," "existentialism" is a term that belongs to intellectual history. Its definition is thus to some extent one of historical convenience. The term was explicitly adopted as a self-description by Jean-Paul Sartre, and through the wide dissemination of the postwar literary and philosophical output of Sartre and his associates—notably Simone de Beauvoir, Maurice Merleau-Ponty, and Albert Camus—existentialism became identified with a cultural movement that flourished in Europe in the 1940s and 1950s. Among the major philosophers identified as existentialists (many of whom—for instance Camus and Heidegger—repudiated the label) were Karl Jaspers, Martin Heidegger, and Martin Buber in Germany, Jean Wahl and Gabriel Marcel in France, the Spaniards José Ortega y Gasset and Miguel de Unamuno, and the Russians Nikolai Berdyaev and Lev Shestov. The nineteenth century philosophers, Søren Kierkegaard and Friedrich Nietzsche, came to be seen as precursors of the movement. Existentialism was as much a literary phenomenon as a philosophical one. Sartre's own ideas were and are better known through his fictional works (such as *Nausea* and *No Exit*) than through his more purely philosophical ones (such as *Being and Nothingness* and *Critique of Dialectical Reason)*, and the postwar years found a very diverse coterie of writers and artists linked under the term: retrospectively, Dostoevsky, Ibsen, and Kafka were conscripted; in Paris there were Jean Genet, André Gide, André Malraux, and the expatriate Samuel Beckett; the Norwegian Knut Hamsun and the Romanian Eugene Ionesco belong to the club; artists such as Alberto Giacometti and even Abstract Expressionists such as Jackson Pollock, Arshile Gorky, and Willem de Kooning, and filmmakers such as Jean-Luc Godard and Ingmar Bergman were understood in existential terms. By the mid 1970s the

cultural image of existentialism had become a cliché, parodized in countless books and films by Woody Allen.

It is sometimes suggested, therefore, that existentialism just is this bygone cultural movement rather than an identifiable philosophical position; or, alternatively, that the term should be restricted to Sartre's philosophy alone. But while a philosophical definition of existentialism may not entirely ignore the cultural fate of the term, and while Sartre's thought must loom large in any account of existentialism, the concept does pick out a distinctive cluster of philosophical problems and helpfully identifies a relatively distinct current of twentieth- and now twenty-first-century philosophical inquiry, one that has had significant impact on fields such as theology (through Rudolf Bultmann, Paul Tillich, Karl Barth, and others) and psychology (from Ludwig Binswanger and Medard Boss to Otto Rank, R. D. Laing, and Viktor Frankl). What makes this current of inquiry distinct is not its concern with "existence" in general, but rather its claim that thinking about *human* existence requires new categories not found in the conceptual repertoire of ancient or modern thought; human beings can be understood neither as substances with fixed properties, nor as subjects interacting with a world of objects.

On the existential view, to understand what a human being is it is not enough to know all the truths that natural science—including the science of psychology—could tell us. The dualist who holds that human beings are composed of independent substances—"mind" and "body"—is no better off in this regard than is the physicalist, who holds that human existence can be adequately explained in terms of the fundamental physical constituents of the universe. Existentialism does not deny the validity of the basic categories of physics, biology, psychology, and the other sciences (categories such as matter, causality, force, function, organism, development, motivation, and so on). It claims only that human beings cannot be fully understood in terms of them. Nor can such an

understanding be gained by supplementing our scientific picture with a *moral* one. Categories of moral theory such as intention, blame, responsibility, character, duty, virtue, and the like *do* capture important aspects of the human condition, but neither moral thinking (governed by the norms of the good and the right) nor scientific thinking (governed by the norm of truth) suffices.

"Existentialism", therefore, may be defined as the philosophical theory which holds that a further set of categories, governed by the norm of *authenticity*, is necessary to grasp human existence. To approach existentialism in this categorial way may seem to conceal what is often taken to be its "heart" (Kaufmann 1968: 12), namely, its character as a gesture of protest against academic philosophy, its anti-system sensibility, its flight from the "iron cage" of reason. But while it is true that the major existential philosophers wrote with a passion and urgency rather uncommon in our own time, and while the idea that philosophy cannot be practiced in the disinterested manner of an objective science is indeed central to existentialism, it is equally true that all the themes popularly associated with existentialism—dread, boredom, alienation, the absurd, freedom, commitment, nothingness, and so on—find their philosophical significance in the context of the search for a new categorial framework, together with its governing norm.

- 1. The Emergence of Existence as a Philosophical Problem
  - 1.1 Kierkegaard: "The Single Individual"
  - 1.2 Nietzsche and Nihilism
- 2. "Existence Precedes Essence"
  - 2.1 Facticity and Transcendence
  - 2.2 Alienation
  - 2.3 Authenticity
- 3. Freedom and Value
  - 3.1 Anxiety, Nothingness, the Absurd

1. The Emergence of Existence as a Philosophical Problem

Sartre's existentialism drew its immediate inspiration from the work of the German philosopher, Martin Heidegger. Heidegger's 1927 *Being and Time*, an inquiry into the "being that we ourselves are" (which he termed "Dasein," a German word for existence), introduced most of the motifs that would characterize later existentialist thinking: the tension between the individual and the "public"; an emphasis on the worldly or "situated" character of human thought and reason; a fascination with liminal experiences of anxiety, death, the "nothing" and nihilism; the rejection of science (and above all, causal explanation) as an adequate framework for understanding human being; and the introduction of "authenticity" as the norm of self-identity, tied to the project of self-definition through freedom, choice, and commitment. Though in 1946 Heidegger would repudiate the retrospective labelling of his earlier work as existentialism, it is in that work that the relevant concept of existence finds its first *systematic* philosophical formulation.[1]

As Sartre and Merleau-Ponty would later do, Heidegger pursued these issues with the somewhat unlikely resources of Edmund Husserl's phenomenological method. And while not all existential philosophers were influenced by phenomenology (for instance Jaspers and Marcel), the philosophical legacy of existentialism is largely tied to the form it took as an existential version of phenomenology. Husserl's efforts in the first decades of the twentieth century had been directed toward establishing a descriptive science of consciousness, by which he understood not the object of the natural science of psychology but the "transcendental" field of intentionality, i.e., that whereby our experience is *meaningful*, an experience of something *as* something. The existentialists welcomed Husserl's doctrine of intentionality as a refutation of the Cartesian view according to which consciousness relates immediately only to its own representations, ideas, sensations. According to Husserl, consciousness is our direct openness to the world, one that is governed categorially (normatively) rather than causally; that is, intentionality is not a property of the individual mind but the categorial framework in which mind and world become intelligible.[2]

A phenomenology of consciousness, then, explores neither the metaphysical composition nor the causal genesis of things, but the "constitution" of their meaning. Husserl employed this method to clarify our experience of nature, the socio-cultural world, logic, and mathematics, but Heidegger argued that he had failed to raise the most fundamental question, that of the "meaning of being" as such. In turning phenomenology toward the question of what it means to be, Heidegger insists that the question be raised *concretely*: it is not at first some academic exercise but a burning concern arising from life itself, the question of what it means for *me* to be. Existential themes take on salience when one sees that the general question of the meaning of being involves first becoming clear about one's own being as an inquirer. According to Heidegger, the categories bequeathed by the philosophical tradition for

understanding a being who can *question* his or her being are insufficient: traditional concepts of a substance decked out with reason, or of a subject blessed with self-consciousness, misconstrue our fundamental character as "being-in-the-world." In his phenomenological pursuit of the categories that govern being-in-the-world, Heidegger became the reluctant father of existentialism because he drew inspiration from two seminal, though in academic circles then relatively unknown, nineteenth-century writers, Søren Kierkegaard and Friedrich Nietzsche. One can find anticipations of existential thought in many places (for instance, in Socratic irony, Augustine, Pascal, or the late Schelling), but the roots of the problem of existence in its contemporary significance lie in the work of Kierkegaard and Nietzsche.

1.1 Kierkegaard: "The Single Individual"

Kierkegaard developed this problem in the context of his radical approach to Christian faith; Nietzsche did so in light of his thesis of the death of God. Subsequent existential thought reflects this difference: while some writers—such as Sartre and Beauvoir—were resolutely atheist in outlook, others—such as Heidegger, Jaspers, Marcel, and Buber—variously explored the implications of the concept "authentic existence" for religious consciousness. Though neither Nietzsche's nor Kierkegaard's thought can be reduced to a single strand, both took an interest in what Kierkegaard termed "the single individual." Both were convinced that this singularity, what is most my own, "me," could be meaningfully reflected upon while yet, precisely because of its singularity, remain invisible to traditional philosophy, with its emphasis either on what follows unerring objective laws of nature or else conforms to the universal standards of moral reason. A focus on existence thus led, in both, to unique textual strategies quite alien to the philosophy of their time.

In Kierkegaard, the singularity of existence comes to light at the moment of conflict between ethics and religious faith. Suppose it is my sense of doing God's will that makes my life meaningful. How does philosophy conceive this meaning? Drawing here on Hegel as emblematic of the entire tradition, Kierkegaard, in his book *Fear and Trembling*, argues that for philosophy my life becomes meaningful when I "raise myself to the universal" by bringing my immediate (natural) desires and inclinations under the moral law, which represents my "telos" or what I *ought* to be. In doing so I *lose* my individuality (since the law holds for all) but my actions become meaningful in the sense of understandable, governed by a norm. Now a person whose sense of doing God's will is what gives her life meaning will be intelligible just to the extent that her action conforms to the universal dictates of ethics. But what if, as in case of Abraham's sacrifice of his son, the action contradicts what ethics demands? Kierkegaard[3] believes *both* that Abraham's life is supremely meaningful (it is not simply a matter of some immediate desire or meaningless tic that overcomes Abraham's ethical consciousness; on the contrary, doing the moral thing is *itself* in this case his tempting inclination) *and* that philosophy cannot understand it, thus condemning it in the name of ethics. God's command here cannot be seen as a law that would pertain to all; it addresses Abraham in his singularity. If Abraham's life is meaningful, it represents, from a philosophical point of view, the "paradox" that through faith the "single individual is higher than the universal." Existence as a philosophical problem appears at this point: if there is a dimension to my being that is both meaningful and yet not governed by the rational standard of morality, by what standard *is* it governed? For unless there is some standard it is idle to speak of "meaning."

To solve this problem there must be a norm inherent in singularity itself, and, in his *Concluding Unscientific Postscript*, Kierkegaard tries to express such a norm in his claim that "subjectivity is the truth," an idea that prefigures the existential concept of authenticity. Abraham has no

objective reason to think that the command he hears comes from God; indeed, based on the content of the command he has every reason, as Kant pointed out in *Religion Within the Limits of Reason Alone*, to think that it *cannot* come from God. His sole justification is what Kierkegaard calls the *passion* of faith. Such faith is, rationally speaking, absurd, a "leap," so if there is to be any talk of truth here it is a standard that measures not the *content* of Abraham's act, but the way in which he accomplishes it. To perform the movement of faith "subjectively" is to embrace the paradox as normative for me in spite of its absurdity, rather than to seek an escape from it by means of objective textual exegesis, historical criticism, or some other strategy for translating the singularity of my situation into the universal. Because my reason cannot help here, the normative appropriation is a function of my "inwardness" or passion. In this way I "truly" become what I nominally already am. To say that subjectivity is the truth is to highlight a way of being, then, and not a mode of knowing; truth measures the attitude ("passion") with which I appropriate, or make my own, an "objective uncertainty" (the voice of God) in a "process of highest inwardness."

In contrast to the singularity of this movement, for Kierkegaard, stands the crowd: "the crowd is untruth." The crowd is, roughly, public opinion in the widest sense—the ideas that a given age takes for granted; the ordinary and accepted way of doing things; the complacent attitude that comes from the conformity necessary for social life—and what condemns it to "untruth" in Kierkegaard's eyes is the way that it insinuates itself into an individual's own sense of who she is, relieving her of the burden of being herself: if everyone is a Christian there is no need for me to "become" one. Since it is a measure not of knowing but of being, one can see how Kierkegaard answers those who object that his concept of subjectivity as truth is based on an equivocation: the objective truths of science and history, however well-established, are in themselves matters of *indifference*; they belong to the crowd. It is not insofar as truth can be

established objectively that it takes on meaning, but rather insofar as it is appropriated "passionately" in its very uncertainty. To "exist" is always to be confronted with this question of meaning. The truths that matter to who one is cannot, like Descartes' *morale definitif*, be something to be attained only when objective science has completed its task.

1.2 Nietzsche and Nihilism

For Kierkegaard existence emerges as a philosophical problem in the struggle to think the paradoxical presence of God; for Nietzsche it is found in the reverberations of the phrase "God is dead," in the challenge of nihilism.

Responding in part to the cultural situation in nineteenth-century Europe —historical scholarship continuing to erode fundamentalist readings of the Bible, the growing cultural capital of the natural sciences, and Darwinism in particular—and in part driven by his own investigations into the psychology and history of moral concepts, Nietzsche sought to draw the consequences of the death of God, the collapse of any theistic support for morality. Like his contemporary, Fyodor Dostoevsky, whose character, Ivan, in *The Brothers Karamazov,* famously argues that if God does not exist then everything is permitted, Nietzsche's overriding concern is to find a way to take the measure of human life in the modern world. Unlike Dostoevsky, however, Nietzsche sees a complicity between morality and the Christian God that perpetuates a life-denying, and so ultimately nihilistic, stance. Nietzsche was not the first to de-couple morality from its divine sanction; psychological theories of the moral sentiments, developed since the eighteenth century, provided a purely human account of moral normativity. But while these earlier theories had been offered as *justifications* of the normative force of morality, Nietzsche's idea that behind moral prescriptions lies nothing but "will to power" undermined that authority. On the account given in *On the Genealogy of Morals*, the

Judeo-Christian moral order arose as an expression of the *ressentiment* of the weak against the power exercised over them by the strong. A tool used to thwart that power, it had over time become internalized in the form of conscience, creating a "sick" animal whose will is at war with its own vital instincts. Thus Nietzsche arrived at Kierkegaard's idea that "the crowd is untruth": the so-called autonomous, self-legislating individual is nothing but a herd animal that has trained itself to docility and unfreedom by conforming to the "universal" standards of morality. The normative is nothing but the normal.

Yet this is not the end of the story for Nietzsche, any more than it was for Kierkegaard. If the autonomous individual has so far signified nothing but herd mentality—if moral norms arose precisely to produce such conformists—the individual nevertheless has the potential to become something else, the sick animal is "pregnant with a future." Nietzsche saw that in the nineteenth century the "highest values" had begun to "devalue themselves." For instance, the Christian value of truth-telling, institutionalized in the form of science, had undermined the belief in God, disenchanting the world and excluding from it any pre-given moral meaning. In such a situation the individual is forced back upon himself. On the one hand, if he is weakly constituted he may fall victim to despair in the face of nihilism, the recognition that life has no instrinsic meaning. On the other hand, for a "strong" or creative individual nihilism presents a liberating opportunity to take responsibility for meaning, to exercise creativity by "transvaluing" her values, establishing a new "order of rank." Through his prophet, Zarathustra, Nietzsche imagined such a person as the "overman" (*Übermensch*), the one who teaches "the meaning of the earth" and has no need of otherworldly supports for the values he embodies. The overman represents a form of life, a mode of existence, that is to blossom from the communalized, moralized "last man" of the nineteenth century. He has understood that nihilism is the ultimate meaning of the moral point

of view, its life-denying essence, and he reconfigures the moral idea of autonomy so as to release the life-affirming potential within it.

Thus, for Nietzsche, existence emerges as a philosophical problem in his distinction between moral autonomy (as obedience to the moral law) and an autonomy "beyond good and evil." But if one is to speak of autonomy, meaning, and value at all, the mode of being beyond good and evil cannot simply be a lawless state of arbitrary and impulsive behavior. If such existence is to be thinkable there must be a standard by which success or failure can be measured. Nietzsche variously indicates such a standard in his references to "health," "strength," and "the meaning of the earth." Perhaps his most instructive indication, however, comes from aesthetics, since its concept of *style*, as elaborated in *The Gay Science*, provides a norm appropriate to the singularity of existence. To say that a work of art has style is to invoke a standard for judging it, but one that cannot be specified in the form of a general law of which the work would be a mere instance. Rather, in a curious way, the norm is internal to the work. For Nietzsche, existence falls under such an imperative of style: to create meaning and value in a world from which all transcendent supports have fallen away is to give unique shape to one's immediate inclinations, drives, and passions; to interpret, prune, and enhance according to a unifying sensibility, a ruling instinct, that brings everything into a whole that satisfies the non-conceptual, aesthetic norm of what fits, what belongs, what is appropriate.

As did Kierkegaard, then, Nietzsche uncovers an aspect of my being that can be understood neither in terms of immediate drives and inclinations nor in terms of a universal law of behavior, an aspect that is measured not in terms of an objective inventory of *what* I am but in terms of my *way* of being it. Neither Kierkegaard nor Nietzsche, however, developed this insight in a fully systematic way. That would be left to their twentieth-century heirs.

2. "Existence Precedes Essence"

Sartre's slogan—"existence precedes essence"—may serve to introduce what is most distinctive of existentialism, namely, the idea that no general, non-formal account of what it means to be human can be given, since that meaning is decided in and through existing itself. Existence is "self-making-in-a-situation" (Fackenheim 1961: 37). In contrast to other entities, whose essential properties are fixed by the *kind* of entities they are, what is essential to a human being—what makes her *who* she is—is not fixed by her type but by what she makes of herself, who she becomes. [4] The fundamental contribution of existential thought lies in the idea that one's identity is constituted neither by nature nor by culture, since to "exist" is precisely to constitute such an identity. It is in light of this idea that key existential notions such as facticity, transcendence (project), alienation, and authenticity must be understood.

At first, it seems hard to understand how one can say much about existence as such. Traditionally, philosophers have connected the concept of existence with that of essence in such a way that the former signifies merely the *instantiation* of the latter. If "essence" designates *what* a thing is and "existence" *that* it is, it follows that what is intelligible about any given thing, what can be thought about it, will belong to its essence. It is from essence in this sense—say, human being as rational animal or *imago Dei*—that ancient philosophy drew its prescriptions for an individual's way of life, its estimation of the meaning and value of existence. Having an essence meant that human beings could be placed within a larger whole, a *kosmos*, that provided the standard for human flourishing. Modern philosophy retained this framework even as it abandoned the idea of a "natural place" for man in the face of the scientific picture of an infinite, labyrinthine universe. In what looks like a proto-existential move, Descartes rejected the traditional essential definitions of man in favor of a radical, first-person reflection on his own existence, the "I am."

Nevertheless, he quickly reinstated the old model by characterizing his existence as that of a substance determined by an essential property, "thinking." In contrast, Heidegger proposes that "I" am "an entity whose what [essence] is precisely to be and nothing but to be" (Heidegger 1985: 110; 1962: 67). Such an entity's existing cannot, therefore, be thought as the instantiation of an essence, and consequently what it means to be such an entity cannot be determined by appeal to pre-given frameworks or systems—whether scientific, historical, or philosophical.

2.1 Facticity and Transcendence

Of course, there is a sense in which human beings do instantiate essences, as Heidegger's phrase already admits.[5] But what matters for existential thought is the *manner* of such instantiation, the *way* of existing. What this means can be seen by contrasting human existence with the modes of being Heidegger terms the "available" (or "ready-to-hand," *zuhanden*) and the "occurrent" (or "present-at-hand," *vorhanden*). Entities of the first sort, exemplified by tools as they present themselves in use, are defined by the social practices in which they are employed, and their properties are established in relation to the norms of those practices. A saw is sharp, for instance, in relation to what counts as successful cutting. Entities of the second sort, exemplified by objects of perceptual contemplation or scientific investigation, are defined by the norms governing perceptual givenness or scientific theory-construction. An available or occurrent entity instantiates some property if that property is truly predicated of it. Human beings can be considered in this way as well. However, in contrast to the previous cases, the fact that natural and social properties can truly be predicated of human beings is not sufficient to determine what it is for *me* to be a human being. This, the existentialists argue, is because such properties are never merely brute determinations of who I am but are always in question. Who I am depends on what I *make* of my "properties"; they matter to me in a way that is impossible for merely available and

occurrent entities. As Heidegger puts it, existence is "care" (*Sorge*): to exist is not simply to be, but to be an *issue* for oneself. In Sartre's terms, while other entities exist "in themselves" (*en soi*) and "are what they are," human reality is also "for itself" (*pour soi*) and thus is not exhausted by any of its determinations. It is what it is not and is not what it is (Sartre 1992: 112).

Human existence, then, cannot be thought through categories appropriate to things: substance, event, process. There is something of an internal distinction in existence that undermines such attempts, a distinction that existential philosophers try to capture in the categories of "facticity" and "transcendence." To be is to co-ordinate these opposed moments in some way, and who I am, my essence, is nothing but my *manner* of co-ordinating them. In this sense human beings make themselves in situation: what I am cannot be separated from what I *take* myself to be. In Charles Taylor's phrase, human beings are "self-interpreting animals" (Taylor 1985: 45), where the interpretation is constitutive of the interpreter. If such a view is not to collapse into contradiction the notions of facticity and transcendence must be elucidated. Risking some oversimplification, they can be approached as the correlates of the two attitudes I can take toward myself: the attitude of third-person theoretical observer and the attitude of first-person practical agent.

Facticity includes all those properties that third-person investigation can establish about me: natural properties such as weight, height, and skin color; social facts such as race, class, and nationality; psychological properties such as my web of belief, desires, and character traits; historical facts such as my past actions, my family background, and my broader historical milieu; and so on.[6] I am not originally aware of my facticity in this third-person way; rather, it is manifest in my moods as a kind of burden, the weight of "having to be." However, I *can* adopt a third-person or objectifying stance toward my own being, and then these aspects of my

facticity may appear precisely as that which defines or determines who I am. From an existential point of view, however, this would be an error —*not* because these aspects of my being are not real or factual, but because the kind of being that *I* am cannot be defined in factual, or third-person, terms.[7] These elements of facticity cannot be said to belong to me in the way that the color of an apple belongs to the apple, for *as* belonging to me, *as* "determining" me, they have always already been interpreted *by* me. Though third-person observation can identify skin color, class, or ethnicity, the minute it seeks to identify them as *mine* it must contend with the distinctive character of the existence I possess. There is no sense in which facticity is both mine *and* merely a matter of fact, since my existence—the kind of being I am—is also defined by the stance I take toward my facticity. This is what existential philosophers call "transcendence."

Transcendence refers to that attitude toward myself characteristic of my practical engagement in the world, the agent's perspective. An agent is oriented by the task at hand as something to be brought about through its own will or agency. Such orientation does not take itself as a theme but loses itself in what is to be done. Thereby, things present themselves not as indifferent givens, facts, but as meaningful: salient, expedient, obstructive, and so on. To speak of "transcendence" here is to indicate that the agent "goes beyond" what simply is toward what can be: the factual—including the agent's own properties—always emerges in light of the possible, where the possible is not a function of anonymous forces (third-person or logical possibility) but a function of the agent's *choice* and *decision*.[8] Just as this suddenly empty pen is either a nettlesome impediment to my finishing this article, or a welcome occasion for doing something else, depending on how I determine my behavior in relation to it, so too my own factic properties—such as irrascibility, laziness, or bourgeois workaholism— take on meaning (become functioning reasons) on the basis of how I endorse or disavow them in the present action.

Existentialists tend to describe the perspective of engaged agency in terms of "choice," and they are sometimes criticized for this. It may be—the argument runs—that I can be said to choose a course of action at the conclusion of a process of deliberation, but there seems to be no choice involved when, in the heat of the moment, I toss the useless pen aside in frustration. Can its being useless be traced back to my "choice" to be frustrated? But the point in using such language is simply to insist that in the first-person perspective of agency I cannot *conceive* myself as determined by anything that is available to me only in third-person terms. Behind the existentialist's insistence that facticity and transcendence remain irreducible aspects of one and the same being is the insight that, for a being who can say "I," the third-person perspective on who one is has no more authority than the first-person (agent's) perspective.[9]

Because existence is co-constituted by facticity and transcendence, the self cannot be conceived as a Cartesian ego but is embodied being-in-the-world, a self-making in situation. It is through transcendence—or what the existentialists also refer to as my "projects"—that the world is revealed, takes on meaning; but such projects are themselves factic or "situated"— not the product of some antecedently constituted "person" or intelligible character but embedded in a world that is decidedly not my representation. Because my projects are who I *am* in the mode of engaged agency (and not like plans that I merely represent to myself in reflective deliberation), the world in a certain sense reveals to me who I am. For reasons to be explored in the next section, the meaning of my choice is not always transparent to me. Nevertheless, because it necessarily reveals the world in a certain way, that meaning, my own "identity," can be discovered by what Sartre calls "existential psychoanalysis." By understanding an individual's patterns of behavior—that is, by reconstructing the meaningful world that such behavior reveals—one can uncover the "fundamental project" or basic choice of oneself that gives distinctive shape to an individual life. Existential psychoanalysis represents a kind of compromise between the

first- and third-person perspectives: like the latter, it objectifies the person and treats its open-ended practical horizons as in a certain sense closed; like the former, however, it seeks to understand the choices from the inside, to grasp the identity of the individual as a matter of the first-person meaning that haunts him, rather than as a function of inert psychic mechanisms with which the individual has no acquaintance.[10]

2.2 Alienation

The anti-Cartesian view of the self as in situation yields the familiar existential theme of the "alienated" self, the estrangement of the self both from the world and from itself. In the first place, though it is through my projects that world takes on meaning, the world itself is not brought into being through my projects; it retains its otherness and thus can come forth as utterly alien, as *unheimlich*. Sometimes translated as "uncanny," this Heideggerian word's stem (*Heim*, "home") points, instead, to the strangeness of a world in which I precisely do *not* feel "at home." (see the section on The Ideality of Values below). This experience, basic to existential thought, contrasts most sharply with the ancient notion of a *kosmos* in which human beings have a well-ordered place, and it connects existential thought tightly to the modern experience of a meaningless universe.

In the second place, the world includes other people, and as a consequence I am not merely the revealer of the world but something revealed in the projects of those others. Thus who I am is not merely a function of my own projects, but is also a matter of my "being-for-others." Sartre (1992: 340-58) brings out this form of alienation in his famous analysis of "the Look." So long as I am engaged unreflectively in a certain practice I am nothing but that first-person perspective which constitutes things as having a distinctive salience in light of what I am doing. I am absorbed in the world and do not experience myself as having an "outside"; that is, I do

not understand my action through some third-person description, as an instance of some general behavior. However, when I become aware of being looked at (that is, when my subjectivity is invaded by the subjectivity of another for whom I am merely part of the world, an item for her projects), I become aware of having a "nature," a "character," of being or doing *something*. I am not merely looking through a keyhole; I am a *voyeur*. I cannot originally experience myself *as* something—a voyeur, for instance. Only the other can give rise to this mode of my being, a mode that I acknowledge as *mine* (and not merely the other's opinion of me) in the *shame* in which I register it. It is because there are others in the world that I can take a third-person perspective on myself; but this reveals the extent to which I am *alienated* from a dimension of my being: who I am in an objective sense can be originally revealed only by the Other. This has implications for existential social theory (see the section on Sartre: Existentialism and Marxism below).

Finally, the self-understanding, or project, thanks to which the world is there for me in a meaningful way, already belongs to that world, derives from it, from the tradition or society in which I find myself. Though it is "me," it is not me "as my own." My very engagement in the world alienates me from my authentic possibility. This theme is brought out most clearly by Heidegger: the anti-Cartesian idea that the self is defined first of all by its practical engagement entails that this self is not properly individual but rather indisinguishable from anyone else (*das Man*) who engages in such practices: such a "they-self" does what "one" does. The idea is something like this: Practices can allow things to show up as meaningful—as hammers, dollar bills, or artworks—because practices involve aims that carry with them norms, satisfaction conditions, for what shows up in them. But norms and rules, as Wittgenstein has shown, are essentially public, and that means that when I engage in practices I must be essentially interchangeable with anyone else who does: I eat as one eats; I drive as one drives; I even protest as one protests. To the extent that

my activity is to be an instance of such a practice, I must do it in the normal way. Deviations can be recognized as deviations only against this norm, and if they deviate too far they can't be recognized at all.[11] Thus, if who I am is defined through existing, this "who" is normally pre-defined by what is average, by the roles available to me in my culture, and so on. The "I" that gets defined is thereby "anonymous," or "anyone"; self-making is largely a function of *not* distinguishing myself from others.

If there is nevertheless good sense in talking of the singularity of my existence, it will not be something with which one starts but something that gets *achieved* in recovering oneself from alienation or lostness in the "crowd." If the normative is first of all the normal, however, it might seem that talk about a norm for the *singularity* of existence, a standard for thinking about what is my *ownmost* just as I myself, would be incoherent. It is here that the idea of "authenticity" must come into focus.

2.3 Authenticity

By what standard are we to think our efforts "to be," our manner of being a self? If such standards traditionally derive from the essence that a particular thing instantiates—this hammer is a good one if it instantiates what a hammer is supposed to be—and if there is nothing that a human being is, by its essence, *supposed* to be, can the meaning of existence at all be thought? Existentialism arises with the collapse of the idea that philosophy can provide substantive norms for existing, ones that specify particular ways of life. Nevertheless, there remains the distinction between what I do "as" myself and as "anyone," so in this sense existing is something at which I can succeed or fail. Authenticity—in German, *Eigentlichkeit*—names that attitude in which I engage in my projects *as* my *own* (*eigen*).

What this means can perhaps be brought out by considering moral evaluations. In keeping my promise I act in accord with duty; and if I keep it *because* it is my duty, I also act morally (according to Kant) because I am acting for the *sake* of duty. But existentially there is still a further evaluation to be made. My moral act is *inauthentic* if, in keeping my promise for the sake of duty, I do so because that is what "one" does (what "moral people" do). But I can do the same thing *authentically* if, in keeping my promise for the sake of duty, acting this way is something I choose *as my own*, something to which, apart from its social sanction, I commit myself. Similarly, doing the right thing from a fixed and stable character — which virtue ethics considers a condition of the good — is not beyond the reach of existential evaluation: such character may simply be a product of my tendency to "do what one does," including feeling "the right way" about things and betaking myself in appropriate ways as one is expected to do. But such character might also be a reflection of my *choice* of myself, a *commitment* I make to be a person of this sort. In both cases I have succeeded in being good; only in the latter case, however, have I succeeded in being *myself*.[12]

Thus the norm of authenticity refers to a kind of "transparency" with regard to my situation, a recognition that I am a being who *can be* responsible for who I am. In choosing in light of this norm I can be said to recover myself from alienation, from my absorption in the anonymous "one-self" that characterizes me in my everyday engagement in the world. Authenticity thus indicates a certain kind of integrity — not that of a pre-given whole, an identity waiting to be discovered, but that of a project to which I can either commit myself (and thus "become" what it entails) or else simply occupy for a time, inauthentically drifting in and out of various affairs. Some writers have taken this notion a step further, arguing that the measure of an authentic life lies in the integrity of a *narrative*, that to be a self is to constitute a story in which a kind of wholeness prevails, to be the author of oneself as a unique individual (Nehamas 1998; Ricoeur

1992). In contrast, the inauthentic life would be one without such integrity, one in which I allow my life-story to be dictated by the world. Be that as it may, it is clear that one can commit oneself to a life of chamealeon-like variety, as does Don Juan in Kierkegaard's version of the legend. Even interpreted narratively, then, the norm of authenticity remains a formal one. As with Kierkegaard's Knight of Faith, one cannot tell who is authentic by looking at the content of their lives.[13]

Authenticity defines a condition on self-making: do I succeed in making *myself*, or will who I am merely be a function of the roles I find myself in? Thus to be authentic can also be thought as a way of being autonomous. In choosing "resolutely"—that is, in commiting myself to a certain course of action, a certain way of being in the world—I have given myself the rule that belongs to the role I come to adopt. The inauthentic person, in contrast, merely *occupies* such a role, and may do so "irresolutely," without commitment. Being a father authentically does not necessarily make me a better father, but what it *means* to be a father has become explicitly *my* concern. It is here that existentialism locates the singularity of existence and identifies what is irreducible in the first-person stance. At the same time, authenticity does not hold out some specific way of life as a norm; that is, it does not distinguish between the projects that I might choose. Instead, it governs the manner in which I am engaged in such projects—either as "my own" or as "what one does," transparently or opaquely.

Thus existentialism's focus on authenticity leads to a distinctive stance toward ethics and value-theory generally. The possibility of authenticity is a mark of my *freedom*, and it is through freedom that existentialism approaches questions of value, leading to many of its most recognizable doctrines.

3. Freedom and Value

Existentialism did not develop much in the way of a normative ethics; however, a certain approach to the theory of value and to moral psychology, deriving from the idea of existence as self-making in situation, are distinctive marks of the existentialist tradition. In value theory, existentialists tend to emphasize the conventionality or groundlessness of values, their "ideality," the fact that they arise entirely through the projects of human beings against the background of an otherwise meaningless and indifferent world. Existential moral psychology emphasizes human freedom and focuses on the sources of mendacity, self-deception, and hypocrisy in moral consciousness. The familiar existential themes of anxiety, nothingness, and the absurd must be understood in this context. At the same time, there is deep concern to foster an authentic stance toward the human, groundless, values without which no project is possible, a concern that gets expressed in the notions of "engagement" and "commitment."[14]

3.1 Anxiety, Nothingness, the Absurd

As a predicate of existence, the concept of freedom is not initially established on the basis of arguments against determinism; nor is it, in Kantian fashion, taken simply as a given of practical self-consciousness. Rather, it is located in the *breakdown* of direct practical activity. The "evidence" of freedom is a matter neither of theoretical nor of practical consciousness but arises from the self-understanding that accompanies a certain *mood* into which I may fall, namely, anxiety (*Angst*, *angoisse*). Both Heidegger and Sartre believe that phenomenological analysis of the kind of intentionality that belongs to moods does not merely register a passing modification of the psyche but reveals fundamental aspects of the self. Fear, for instance, reveals some region of the world as threatening, some element in it as a threat, and myself as vulnerable. In anxiety, as in fear, I grasp myself as threatened or as vulnerable; but unlike fear, anxiety has no direct object, there is nothing in the world that is threatening. This

is because anxiety pulls me altogether out of the circuit of those projects thanks to which things are there for me in meaningful ways; I can no longer "gear into" the world. And with this collapse of my practical immersion in roles and projects, I also lose the basic sense of who I am that is provided by these roles. In thus robbing me of the possibility of practical self-identification, anxiety teaches me that I do not coincide with anything that I factically am. Further, since the identity bound up with such roles and practices is always typical and public, the collapse of this identity reveals an ultimately first-personal aspect of myself that is irreducible to *das Man*. As Heidegger puts it, anxiety testifies to a kind of "existential solipsism." It is this reluctant, because disorienting and dispossessing, retreat into myself in anxiety that yields the existential figure of the outsider, the isolated one who "sees through" the phoniness of those who, unaware of what the breakdown of anxiety portends, live their lives complacently identifying with their roles as though these roles thoroughly defined them. While this "outsider" stance may be easy to ridicule as adolescent self-absorption, it is also solidly supported by the phenomenology (or moral psychology) of first-person experience.

The experience of anxiety also yields the existential theme of the *absurd*, a version of what was previously introduced as alienation from the world (see the section on Alienation above). So long as I am gearing into the world practically, in a seamless and absorbed way, things present themselves as meaningfully co-ordinated with the projects in which I am engaged; they show me the face that is relevant to what I am doing. But the connection between these meanings and my projects is not itself something that I experience. Rather, the hammer's usefulness, its value as a hammer, appears simply to belong to it in the same way that its weight or color does. So long as I am practically engaged, in short, all things appear to have reasons for being, and I, correlatively, experience myself as fully at home in the world. The world has an order that is largely transparent to me (even its mysteries are grasped simply as something for

which there are reasons that are there "for others," for "experts," merely beyond my limited horizon). In the mood of anxiety, however, it is just this character that fades from the world. *Because* I am no longer practically engaged, the meaning that had previously inhabited the thing as the density of its being now stares back at me as a mere *name*, as something I "know" but which no longer claims me. As when one repeats a word until it loses meaning, anxiety undermines the taken-for-granted sense of things. They become *absurd*. Things do not disappear, but all that remains of them is the blank recognition *that* they are—an experience that informs a central scene in Sartre's novel *Nausea*. As Roquentin sits in a park, the root of a tree loses its character of familiarity until he is overcome by nausea at its utterly alien character, its being *en soi*. While such an experience is no *more* genuine than my practical, engaged experience of a world of meaning, it is no *less* genuine either. An existential account of meaning and value must recognize *both* possibilities (and their intermediaries). To do so is to acknowledge a certain absurdity to existence: though reason and value have a foothold in the world (they are not, after all, my arbitrary invention), they nevertheless lack any ultimate foundation. Values are not intrinsic to being, and at some point reasons give out.[15]

Another term for the groundlessness of the world of meaning is "nothingness." Heidegger introduced this term to indicate the kind of self- and world-understanding that emerges in anxiety: because my practical identity is constituted by the practices I engage in, when these collapse I "am" not anything. In a manner of speaking I am thus brought face-to-face with my own finitude, my "death" as the possibility in which I am no longer able to *be* anything. This experience of my own death, or "nothingness," in anxiety can act as a spur to authenticity: I come to see that I "am" not anything but must "make myself be" through my choice. In *commiting* myself in the face of death—that is, aware of the nothingness of my identity if not supported by me right up to the end—the roles that I

have hitherto thoughtlessly engaged in as one does now become something that I myself own up to, become responsible for. Heidegger termed this mode of self-awareness—awareness of the ultimate nothingness of my practical identity—"freedom," and Sartre developed this existential concept of freedom in rich detail. This is not to say that Heidegger's and Sartre's views on freedom are identical. Heidegger, for instance, will emphasize that freedom is always "thrown" into an historical situation from which it draws its possibilities, while Sartre (who is equally aware of the "facticity" of our choices) will emphasize that such "possibilities" nevertheless underdetermine choice. But the theory of radical freedom that Sartre develops is nevertheless directly rooted in Heidegger's account of the nothingness of my practical identity.

Sartre (1992: 70) argues that anxiety provides a lucid experience of that freedom which, though often concealed, characterizes human existence as such. For him, freedom is the dislocation of consciousness from its object, the fundamental "nihilation" or negation by means of which consciousness can grasp its object without losing itself in it: to be conscious of something is to be conscious of *not* being it, a "not" that arises in the very structure of consciousness as being for-itself. Because "nothingness" (or nihilation) is just what consciousness is, there can be no objects *in* consciousness, but only objects *for* consciousness.[16] This means that consciousness is radically free, since its structure precludes that it either *contain* or be *acted on* by things. For instance, because it is not thing-like, consciousness is free with regard to its own prior states. Motives, instincts, psychic forces, and the like cannot be understood as inhabitants of consciousness that might infect freedom from within, inducing one to act in ways for which one is not responsible; rather, they can exist only *for* consciousness as matters of choice. I must either reject their claims or avow them. For Sartre, the ontological freedom of existence entails that determinism is an *excuse* before it is a theory: though through its structure of nihilation consciousness escapes that which would define it—including its own past

choices and behavior—there are times when I may wish to deny my freedom. Thus I may attempt to constitute these aspects of my being as objective "forces" which hold sway over me in the manner of relations between things. This is to adopt the third-person stance on myself, in which what is originally structured in terms of freedom appears as a causal property of myself. I can try to look upon myself as the Other does, but as an excuse this flight from freedom is shown to fail, according to Sartre, in the experience of *anguish*.

For instance, Sartre writes of a gambler who, after losing all and fearing for himself and his family, retreats to the reflective behavior of resolving never to gamble again. This motive thus enters into his facticity as a choice he has made; and, as long as he retains his fear, his living sense of himself as being threatened, it may appear to him that this resolve actually has causal force in keeping him from gambling. However, one evening he confronts the gaming tables and is overcome with anguish at the recognition that his resolve, while still "there," retains none of its power: it is an object *for* consciousness but is not (and never could have been) something *in* consciousness that was determining his actions. In order for it to influence his behavior he has to avow it afresh, but this is just what he cannot do; indeed, just this is what he hoped the original resolve would spare him from having to do. He will have to "remake" the self who was in the original situation of fear and threat. At this point, perhaps, he will try to relieve himself of freedom by giving in to the urge to gamble and chalking it up to "deeper" motives that overcame the initial resolve, problems from his childhood perhaps. But anguish can recur with regard to this strategy as well—for instance, if he needs a loan to continue gambling and must convince someone that he is "as good as his word." The possibilities for self-deception in such cases are endless.

As Sartre points out in great detail, anguish, as the consciousness of freedom, is not something that human beings welcome; rather, we seek

stability, identity, and adopt the language of freedom only when it suits us: those acts are considered by me to be my free acts which exactly match the self I want others to take me to be. We are "condemned to be free," which means that we can never simply *be* who we are but are separated from ourselves by the nothingness of having perpetually to re-choose, or re-commit, ourselves to what we do. Characteristic of the existentialist outlook is the idea that we spend much of lives devising strategies for denying or evading the anguish of freedom. One of these strategies is "bad faith." Another is the appeal to values.

3.2 The Ideality of Values

The idea that freedom is the origin of value—where freedom is defined not in terms of acting rationally (Kant) but rather existentially, as choice and transcendence—is the idea perhaps most closely associated with existentialism. So influential was this general outlook on value that Karl-Otto Apel (1973: 235) came to speak of a kind of "official complementarity of existentialism and scientism" in Western philosophy, according to which what can be justified rationally falls under the "value-free objectivism of science" while all other validity claims become matters for an "existential subjectivism of religious faith and ethical decisions." Positivism attempted to provide a theory of "cognitive meaning" based on what it took to be the inner logic of scientific thought, and it relegated questions of value to cognitive meaninglessness, reducing them to issues of emotive response and subjective preference. While it does not explain evaluative language solely as a function of affective attitudes, existential thought, like positivism, denies that values can be grounded in being—that is, that they can become the theme of a scientific investigation capable of distinguishing true (or valid) from false values.[17] In this regard Sartre speaks of the "ideality" of values, by which he means *not* that they have some sort of timeless validity but that they have no real authority and cannot be used to underwrite or justify my behavior. For Sartre, "values

derive their meaning from an original projection of myself which stands as my choice of myself in the world." But if that is so, then I cannot, without circularity, appeal to values in order to justify this very choice: "I make my decision concerning them—without justification and without excuse" (Sartre 1992: 78). This so-called "decisionism" has been a hotly contested legacy of existentialism and deserves a closer look here.

How is it that values are supposed to be grounded in freedom? By "value" Sartre means those aspects of my experience that do not merely causally effectuate something but rather make a *claim* on me: I do not just see the homeless person but encounter him as "to be helped"; I do not just hear the other's voice but register "a question to be answered honestly"; I do not simply happen to sit quietly in Church but "attend reverently"; I do not merely hear the alarm clock but am "summoned to get up." Values, then, as Sartre writes, appear with the character of *demands* and as such they "lay claim to a foundation" or justification (Sartre 1992: 76). Why *ought* I help the homeless, answer honestly, sit reverently, or get up? Sartre does not claim that there is no answer to these questions but only that the answer depends, finally, on my choice of "myself" which cannot in turn be justfied by appeal to a value. As he puts it, "value derives its being from its exigency and not its exigency from its being." The exigency of value cannot be grounded in being itself, since it would thereby lose its character as an ought; it would "cease even to be value" since it would have the kind of exigency (contrary to freedom) possessed by a mere *cause*. Thus, against then-current value-theoretical intuitionism, Sartre denies that value can "deliver itself to a contemplative intuition which would apprehend it as *being* value and thereby would derive from it its right over my freedom." Instead, "it can be revealed only to an *active* freedom which makes it exist as a value by the sole fact of *recognizing* it as such" (Sartre 1992: 76).

For instance, I do not grasp the exigency of the alarm clock (its character as a demand) in a kind of disinterested perception but only in the very act of responding to it, of getting up. If I fail to get up the alarm has, to that very extent, lost its exigency. Why *must* I get up? At this point I may attempt to justify its demand by appeal to other elements of the situation with which the alarm is bound up: I must get up because I must go to work. From this point of view the alarm's demand appears—and *is*—justified, and such justification will often suffice to get me going again. But the question of the foundation of value has simply been displaced: now it is my job that, in my active engagement, takes on the unquestioned exigency of a demand or value. But it too derives its being as a value from its exigency—that is, from my unreflective engagement in the overall practice of going to work. *Ought* I go to work? Why not be "irresponsible"? If a man's got to eat, why not rather take up a life of crime? If these questions have answers that are themselves exigent it can only be because, at a still deeper level, I am engaged as having chosen myself as a person of a certain sort: respectable, responsible. From *within* that choice there is an answer of what I ought to do, but outside that choice there is none—why should I be respectable, law-abiding?—for it is only because *some* choice has been made that anything at all can appear as compelling, as making a *claim* on me. Only if I am at some level *engaged* do values (and so justification in terms of them) appear at all. The more I pull out of engagement toward reflection on and questioning of my situation, the more I am threatened by ethical anguish—"which is the recognition of the ideality of values" (Sartre 1992: 76). And, as with all anguish, I do not escape this situation by discovering the true order of values but by plunging back into action. If the idea that values are without foundation in being can be understood as a form of nihilism, the existential response to this condition of the modern world is to point out that meaning, value, is not first of all a matter of contemplative theory but a consequence of engagement and commitment.

Thus value judgments can be justified, but only relative to some concrete and specific project. The "pattern of behavior" of the typical bourgeois defines the meaning of "respectability" (Sartre 1992: 77), and so it is true of some particular bit of behavior that it is either respectable or not. For this reason I can be in error about what I ought to do. It may be that something that appears exigent during the course of my unreflective engagement in the world is something that I ought not to give in to. If, thanks to my commitment to the Resistance, a given official appears to me as to be shot, I might nevertheless be wrong to shoot him—if, for instance, the official was not who I thought he was, or if killing him would in fact prove counter-productive given my longer-term goals. Sartre's fictional works are full of explorations of moral psychology of this sort. But I cannot extend these "hypothetical" justifications to a point where some purely theoretical consideration of my obligations—whether derived from the will of God, from Reason, or from the situation itself—could underwrite my freedom in such a way as to relieve it of responsibility. For in order for such considerations to *count* I would have to make myself the sort of person for whom God's will, abstract Reason, or the current situation is *decisive*. For existentialists like Sartre, then, I am "the one who finally makes values exist in order to determine [my] actions by their demands."[18]

Commitment—or "engagement"—is thus ultimately the basis for an authentically meaningful life, that is, one that answers to the existential condition of being human and does not flee that condition by appeal to an abstract system of reason or divine will. Yet though I alone can commit myself to some way of life, some project, I am never alone when I do so; nor do I do so in a social, historical, or political vaccuum. If transcendence represents my radical freedom to define myself, facticity—that other aspect of my being—represents the *situated* character of this self-making. Because freedom as transcendence undermines the idea of a stable, timeless system of moral norms, it is little wonder that existential

philosophers (with the exception of Simone de Beauvoir) devoted scant energy to questions of normative moral theory. However, because this freedom is always socially (and thereby historically) situated, it is equally unsurprising that their writings are greatly concerned with how our choices and commitments are concretely contextualized in terms of political struggles and historical reality.

4. Politics, History, Engagement

For the existentialists engagement is the source of meaning and value; in choosing myself I in a certain sense make my world. On the other hand, I always choose myself in a context where there are others doing the same thing, and in a world that has always already been there. In short, my acting is situated, both socially and historically. Thus, in choosing myself in the first-person singular, I am also choosing in such a way that a first-person plural, a "we," is simultaneously constituted. Such choices make up the domain of social reality: they fit into a pre-determined context of roles and practices that go largely unquestioned and may be thought of as a kind of collective identity. In social action my identity takes shape against a background (the collective identity of the social formation) that remains fixed. On the other hand, it can happen that my choice puts this social formation or collective identity itself into question: who I am to be is thus inseperable from the question of who *we* are to be. Here the first-person plural is itself the issue, and the action that results from such choices constitutes the field of the *political*.

If authenticity is the category by which I am able to think about what it means to "exist," then, the account of authenticity cannot neglect the social, historical, and political aspects of that existence. Thus it is not merely because twentieth-century existentialism flourished at a time when European history appeared to collapse and political affairs loomed especially large that existential philosophers devoted much attention to

these matters; rather, the demand for an account of the "situation" stems from the very character of existence itself, which, unlike the classical "rational subject," is what it is only in relation to its "time." This is not to say, however, that existential philosophers are unanimous in their account of the importance of historical factors or in their estimation of the political in relation to other aspects of existence. Emmanuel Levinas, for example, whose early work belonged within the orbit of existential philosophy, opposed to the "horizontal" temporality of political history a "vertical" or eschatological temporality that radically challenged all historical meaning, while Sartre, in contrast, produced a version of Marxist historical materialism in which existentialism itself became a mere ideology. But we cannot stop to examine all such differences here. Instead, we shall look at the positions of Heidegger and Sartre, who provide opposing examples of how an authentic relation to history and politics can be understood.

4.1 Heidegger: History as Claim

For Heidegger, to exist is to be historical. This does not mean that one simply finds oneself at a particular moment in history, conceived as a linear series of events. Rather, it means that selfhood has a peculiar temporal structure that is the *origin* of that "history" which subsequently comes to be narrated in terms of a series of events. Existential temporality is not a sequence of instants but instead a unified structure in which the "future" (that is, the possibility aimed at in my project) recollects the "past" (that is, what no longer needs to be done, the completed) so as to give meaning to the "present" (that is, the things that take on significance in light of what currently needs doing). To act, therefore, is, in Heidegger's terms, to "historize" (*geschehen*), to constitute something like a narrative unity, with beginning, middle, and end, that does not so much take place *in* time as provides the *condition* for linear time. To exist "between birth and death," then, is not merely to be present in each of a discrete series of temporal instants but to consitute oneself in the unity of a history, and

authentic existence is thus one in which the projects that give shape to existence are ones to which I commit myself in light of this history. Though it belongs to, and defines, a "moment," choice cannot be simply "of the moment"; to be authentic I must understand my choice in light of the potential *wholeness* of my existence.[19]

That this choice has a political dimension stems from the fact that existence is always being-with-others. Though authenticity arises on the basis of my being alienated, in anxiety, from the claims made by norms belonging to the everyday life of *das Man*, any concrete commitment that I make in the movement to recover myself will enlist those norms in two ways. First, *what* I commit myself to will always be derived from (though not reducible to) some "possibility of Dasein that has been there" (Heidegger 1962: 438): I cannot make my identity from whole cloth; I will always understand myself in terms of some way of existing that has been handed down within my tradition.[20] I "choose my hero" (Heidegger 1962: 437) by, for instance, committing myself to a philosophical life, which I understand on the model of Socrates, or to a religious life, which I understand on the model of St. Francis. The point is that I must understand myself in terms of *something*, and these possibilities for understanding come from the historical heritage and the norms that belong to it. Heidegger thinks of this historical dimension as a kind of "fate" (*Schicksal*): not something inevitable that controls my choice but something that, inherited from my historical situation, *claims* me, holds a kind of authority for me.

The second way in which the everyday norms of *das Man* are enlisted in authentic choice stems from the fact that when I commit myself to my "fate" I do so "in and with my 'generation'" (Heidegger 1962: 436). The idea here seems roughly to be this: To opt for a way of going on is to affirm the norms that belong to it; and because of the nature of normativity it is not possible to affirm norms that would hold *only* for *me*. There is a

kind of publicity and scope in the normative such that, when I choose, I exemplify a standard for others as well. Similarly, Heidegger holds that the sociality of my historizing restricts what can be a genuine "fate" or choice for me. Acting is always with others—more specifically, with a "community" or a "people" (*Volk*)—and together this "co-historizing" responds to a "destiny" (*Geschick*) which has guided our fates in advance (Heidegger 1962: 436). Not everything is really possible for us, and an authentic choice must strive to respond to the claim that history makes on the people with whom one belongs, to seize its "destiny." Along this communitarian axis, then, existential historicality can open out onto the question of politics: who are "we" to be?

Heidegger suggests that it was this concept of historicality that underwrote his own concrete political engagement during the period of National Socialism in Germany. Disgusted with the political situation in Weimar Germany and characterizing it as especially irresolute or inauthentic, Heidegger looked upon Hitler's movement as a way of recalling the German people back to their "ownmost" possibility—i.e., a way for Germany to constitute itself authentically as an alternative to the political models of Russia and the United States. Heidegger's choice to intervene in university politics at this time was thus both a choice of himself—in which he chose his hero: Plato's "philosopher-king" (see Arendt 1978)—and a choice for his "generation." Much is controversial about Heidegger's engagement for National Socialism (not least whether he drew the appropriate consequences from his own concept of authenticity), but it provides a clear example of a kind of existential politics that depends on an ability to "tell time"—that is, to sense the imperatives of one's factic historical situation. Heidegger later became very suspicious of this sort of existential politics. Indeed, for the idea of authenticity as resolute commitment he substituted the idea of a "releasement" (*Gelassenheit*) and for engagement the stance of "waiting." He came to believe that the problems that face us (notably, the dominance of technological ways of

thinking) have roots that lie deeper than can be addressed through politics directly. He thus famously denied that democracy was sufficient to deal with the political crisis posed by technology, asserting that "only a god can save us" (Heidegger 1981: 55, 57). But even here, in keeping with the existential notion of historicity, Heidegger's recommendations turn on a reading of history, of the meaning of our time.

4.2 Sartre: Existentialism and Marxism

A very different reading, and a very different recommendation, can be found in the work of Sartre. The basis for Sartre's reading of history, and his politics, was laid in that section of *Being and Nothingness* that describes the birth of the social in the "Look"(*le regard*) of the other. In making me an object for his projects, the other alienates me from myself, displaces me from the subject position (the position from which the world is defined in its meaning and value) and constitutes me *as* something. Concretely, what I am constituted "as" is a function of the other's project and not something that I can make myself be. I am constituted as a "Frenchman" in and through the hostility emanating from that German; I am constituted as a "man" in the resentment of that woman; I am constituted as a "Jew" on the basis of the other's anti-semitism; and so on. This sets up a dimension of my being that I can neither control nor disavow, and my only recourse is to wrench myself away from the other in an attempt to restore myself to the subject-position. For this reason, on Sartre's model, social reality is in perpetual conflict—an Hegelian dialectic in which, for ontological reasons, no state of mutual recognition can ever be achieved. The "we"—the political subject—is always contested, conflicted, unstable.

But this instability does have a certain structure, one which Sartre, steeped in the Marxism of inter-war French thought (Alexandre Kojève, Jean Hyppolite), explored in terms of a certain historical materialism. For social

relations take place not only between human beings but also within institutions that have developed historically and that enshrine relations of power and domination. Thus the struggle for who will take the subject position is not carried out on equal terms. As Simone de Beauvoir demonstrated in detail in her book, *The Second Sex*, the historical and institutional place of women is defined in such a way that they are consigned to a kind of permanent "object" status—they are the "second" sex since social norms are defined in male terms. This being so, a woman's struggle to develop self-defining projects is constrained by a permanent institutional "Look" that already defines her as "woman," whereas a man need not operate under constraints of gender; he feels himself to be simply "human," pure subjectivity. Employing similar insights in reflection on the situations of ethnic and economic oppression, Sartre sought a way to derive political imperatives in the face of the groundlessness of moral values entailed by his view of the ideality of values.

At first, Sartre argued that there was one value—namely freedom itself— that *did* have a kind of universal authority. To commit oneself to anything is also always to commit oneself to the value of freedom. In "Existentialism is a Humanism" Sartre tried to establish this by way of a kind of transcendental argument, but he soon gave up that strategy and pursued the more modest one of claiming that the *writer* must always engage "on the side of freedom." According to the theory of "engaged literature" expounded in *What is Literature?*, in creating a literary world the author is always acting either to imagine paths toward overcoming concrete unfreedoms such as racism and capitalist exploitation, or else closing them off. In the latter case, he is contradicting himself, since the very idea of writing presupposes the freedom of the reader, and that means, in principle, the whole of the reading public. Whatever the merits of this argument, it does suggest the political value to which Sartre remained committed throughout his life: the value of freedom as self-making.

This commitment finally led Sartre to hold that existentialism itself was only an "ideological" moment within Marxism, which he termed "the one philosophy of our time which we cannot go beyond" (Sartre 1968:xxxiv). As this statement suggests, Sartre's embrace of Marxism was a function of his sense of history as the factic situation in which the project of self-making takes place. Because existing is self-making (action), philosophy—including existential philosophy—cannot be understood as a disinterested theorizing about timeless essences but is always already a form of engagement, a diagnosis of the past and a projection of norms appropriate to a different future in light of which the present takes on significance. It therefore always arises from the historical-political situation and is a way of intervening in it. Marxism, like existentialism, makes this necessarily practical orientation of philosophy explicit.

From the beginning existentialism saw itself in this activist way (and this provided the basis for the most serious disagreements among French existentialists such as Sartre, Merleau-Ponty, and Camus, many of which were fought out in the pages of the journal founded by Sartre and Merleau-Ponty, *Les Temps Modernes*). But the later Sartre came to hold that a philosophy of self-making could not content itself with highlighting the situation of individual choice; an authentic *political* identity could only emerge from a theory that situated such choice in a practically oriented analysis of its concrete situation. Thus it appeard to him that the "ideology of existence" was itself merely an alienated form of the deeper analysis of social and historical reality provided by Marx's dialectical approach. In focusing on the most important aspects of the material condition in which the existential project of self-making takes place—namely, economic relations under conditions of scarcity—Marx's critique of capital offered a set of considerations that no "philosophy of freedom" could ignore, considerations that would serve to orient political engagement until such time as "there will exist for *everyone* a margin of *real* freedom beyond the production of life" (Sartre 1968: 34). Marxism is unsurpassable, therefore,

because it is the most lucid theory of our alienated situation of concrete unfreedom, oriented toward the practical-political overcoming of that unfreedom.

Sartre's relation to orthodox Marxism was marked by tension, however, since he held that existing Marxism had abandoned the promise of its dialectical approach to social reality in favor of a dogmatic "apriorism" that subsumed historical reality under a blanket of lifeless abstractions. He thus undertook his *Critique of Dialectical Reason* to restore the promise of Marxism by reconceiving its concept of *praxis* in terms of the existential notion of *project*. What had become a rigid economic determinism would be restored to dialectical fluidity by recalling the existential doctrine of self-making: it is true that man is "made" by history, but at the same time he is making that very history. This attempt to "reconquer man within Marxism" (Sartre 1968: 83)—i.e., to develop a method which would preserve the concrete details of human reality as *lived experience*—was not well received by orthodox Marxists. Sartre's fascination with the details of Flaubert's life, or the life of Baudelaire, smacked too much of "bourgeois idealism." But we see here how Sartre's politics, like Heidegger's, derived from his concept of history: there are no iron-clad laws that make the overthrow of capitalism the inevitable outcome of economic forces; there are only men in situation who make history as they are made by it. Dialectical materialism is the unsurpassable philosophy of those who choose, who commit themselves to, the value of freedom. The political claim that Marxism has on us, then, would rest upon the ideological enclave within it: authentic existence as choice.

Authentic existence thus has an historical, political dimension; all choice will be attentive to history in the sense of contextualizing itself in some temporally narrative understanding of its place. But even here it must be admitted that what makes existence authentic is not the correctness of the narrative understanding it adopts. Authenticity does not depend on some

particular substantive view of history, some particular theory or empirical story. From this point of view, the substantive histories adopted by existential thinkers as different as Heidegger and Sartre should perhaps be read less as scientific accounts, defensible in third-person terms, than as articulations of the historical situation from the perspective of what that situation is taken to demand, given the engaged commitment of their authors. They stand, in other words, less as *justifications* for their authors' existential and political commitments than as themselves a *form* of politics: invitations to others to see things as the author sees them, so that the author's commitment to going on in a certain way will come to be shared.

5. Existentialism Today

As a cultural movement, existentialism belongs to the past. As a philosophical inquiry that introduced a new norm, authenticity, for understanding what it means to be human—a norm tied to a distinctive, post-Cartesian concept of the self as practical, embodied, being-in-the-world—existentialism has continued to play an important role in contemporary thought in both the continental and analytic traditions. The Society for Phenomenology and Existential Philosophy, as well as societies devoted to Heidegger, Sartre, Merleau-Ponty, Jaspers, Beauvoir, and other existential philosophers, provide a forum for ongoing work—both of a historical, scholarly nature and of more systematic focus—that derives from classical existentialism, often bringing it into confrontation with more recent movements such as structuralism, deconstruction, hermeneutics, and feminism. In the area of gender studies Judith Butler (1990) draws importantly on existential sources, as does Lewis Gordon (1995) in the area of race theory (see also Bernasconi 2003). Matthew Ratcliffe (2008) develops an existential approach to psychopathology.

Interest in a narrative conception of self-identity—for instance, in the work of Charles Taylor (1999), Paul Ricoeur, David Carr (1986), or Charles Guignon—has its roots in the existential revision of Hegelian notions of temporality and its critique of rationalism. Hubert Dreyfus (1979) developed an influential criticism of the Artificial Intelligence program drawing essentially upon the existentialist idea, found especially in Heidegger and Merleau-Ponty, that the human world, the world of meaning, must be understood first of all as a function of our embodied practices and cannot be represented as a logically structured system of representations. Calling for a "new existentialism," John Haugeland (1998) has explored the role of existential commitment in scientific practices as truth-tracking practices. In a series of books, Michael Gelven (e.g., 1990, 1997) has reflected upon the distinctions between existential, moral, and epistemological or logical dimensions of experience, showing how the standards appropriate to each intertwine, without reducing to any single one. A revival of interest in moral psychology finds many writers who are taking up the question of self-identity and responsibility in ways that recall the existential themes of self-making and choice—for instance, Christine Korsgaard (1996) appeals crucially to notions of "self-constitution" and "practical identity"; Richard Moran (2001) emphasizes the connection between self-avowal and the first-person perspective in a way that derives in part from Sartre; and Thomas Nagel has followed the existentialist line in connecting meaning to the consciousness of death. Even if such writers tend to proceed with more confidence in the touchstone of rationality than did the classical existentialists, their work operates on the terrain opened up by the earlier thinkers.

In addition, after years of being out of fashion in France, existential motifs have once again become prominent in the work of leading thinkers. Foucault's embrace of a certain concept of freedom, and his exploration of the "care of the self," recall debates within existentialism, as does Derrida's recent work on religion without God and his reflections on the

concepts of death, choice, and responsibility. In very different ways, the books by Cooper (1999) and Alan Schrift (1995) suggest that a re-appraisal of the legacy of existentialism is an important agenda item of contemporary philosophy. Reynolds (2006), for instance, concludes his introduction to existentialism with a consideration of how post-structuralists such as Derrida, Deleuze, and Foucault extend certain reflections found in Sartre, Camus, and Heidegger, while Reynolds (2004) does the same, in more detail for Derrida and Merleau-Ponty. If existentialism's very notoriety as a cultural movement may have impeded its serious philosophical reception, then, it may be that what we have most to learn from existentialism still lies before us.

There are, in fact, reasons to think that such a re-evaluation is currently underway. Several publications that have appeared since the last revision of this article (2010) take up the challenge of bringing existential thought into dialogue with items on the contemporary philosophical agenda. Edward Baring (2011) exhumes the historical relation between Derrida and existentialism and finds a kind of "'Christian' existentialism" in Derrida's work prior to 1952, traces of which are discernible in his later thinking. The collection edited by Judaken and Bernasconi (2012) explores the historical context of existentialist writings informed by contemporary critiques of canonization, while Margaret Simons (2013) re-evaluates the role of Beauvoir, and of feminist thought, in the origins of existentialism itself. In 2011 *The Continuum Companion to Existentialism* appeared (Joseph, Reynolds, and Woodward 2011), followed by *The Cambridge Companion to Existentialism* (Crowell 2012). Articles in both volumes are committed to showing the systematic relevance of existential concepts and approaches for contemporary work in philosophy and other fields. Finally, Aho (2014) highlights how, in areas as diverse as cognitive science, psychiatry, health care, and environmental philosophy, "the legacy of existentialism is alive and well" (2014: 140).