CHAPTER SIX

Putting concepts under stress (I)

6.1 The limits on science

Aristotle was a physicist and a biologist; Francis Bacon, a physicist; Descartes, Newton, and Leibniz also were physicists; so was Immanuel Kant; Charles S. Peirce was a physicist, astronomer, and geodesist; William James was a psychologist; Ludwig Wittgenstein, an aeronautical engineer; Pierre Teilhard de Chardin, a paleontologist; and Noam Chomsky, a linguist. The list of philosophers who were also scientists goes on and on. Indeed, until the nineteenth century there was little distinction between philosophy per se and science. Both were concerned with plumbing the secrets of nature. Departments of physics in universities were usually called “natural philosophy”. Scientific journals often incorporated the word “philosophy” in their titles. One modern physics journal retains its title of 1798: *Philosophical Magazine*.[1] The oldest (1744) scientific society in the United States, the American Philosophical Society, still bears the name given it by its founder, Benjamin Franklin ([23], 5-9). Only in the nineteenth century, particularly with the growing emphasis on the distinction between a priori knowledge and empirical knowledge, did the partnership of two thousand years between philosophy and science begin to become undone. Within the universities, physics departments and psychology departments split off from philosophy departments in order to practice ‘empirical’ science, leaving philosophy – it was

[1] *Philosophical Magazine* is published each month, in three volumes, “A”, “B”, and “Letters”. The journal is devoted to topics “in the experimental, theoretical and applied physics of condensed matter. Part A deals with Defects and Mechanical Properties; Part B with Structural, Electronic, Optical and Magnetic Properties.” (See “Notes for Contributors”, vol. 58A, no. 6 [Dec. 1988].)

imagined – to pursue knowledge in the a priori manner of mathematics and logic.[2]

The split was a historical aberration. It came about through the naive belief that natural science could be pursued free of philosophical ideologies or metaphysical world-views. But there never has been a time when science was essentially and solely empirical and there never has been a time when philosophy was essentially and solely a priori. The difference between the practice of science and of philosophy is, and always has been, one of a matter of the degree to which researchers conduct empirical research. No science is ever totally free of philosophical components, and little, if any, philosophy ever springs out of the resources of pure reason uninformed by any empirical data whatsoever. Even if few philosophers themselves conduct laboratory experiments, they must nonetheless take cognizance of the results of experimental research in their own philosophical pursuits.

Because it was prompted by mistaken views about the very nature of science and of philosophy, the artificial parting of philosophy and science in the nineteenth century never could be very thorough. Too many problems in the sciences – the record is virtually endless – have since cropped up which positively demand philosophical examination: the advent of non-Euclidean geometries; the demise of the absolute theories of space and time; the attack on the Newtonian world-view; the rise of evolutionary theory; the unleashing of nuclear destruction; the appearance of artificial intelligence; the discovery of split-brain phenomena; the challenge of the possibility of paranormal experiences; the technological ability to extend bodily functions past brain death; etc. Bare empirical science is inadequate to provide a human context, a sophisticated understanding, of the implications and relevance of such a flood and diversity of information. The techniques of empirical research do not provide the scientist with the conceptual tools needed to synthesize a satisfactory, comprehensive world-view out of these disparate pieces.

[2] Some philosophy departments were to experience successive, if not exactly systematic, fragmentation. At Harvard, for example, the Department of Political Economy broke away from the Philosophy Department in 1879; the Department of Social Ethics, in 1906; and the Division of Education, also in 1906 ([38]). The Psychology Department, however, unlike those at most other American Universities where the split had come earlier, remained as a division within Philosophy until 1934 ([24], 24, and [118], 459-63).

If science cannot cope without philosophy, the latter, in its turn, withers without the stimulus of science. Ethics could never be quite the same once travelers brought back news of the incredible diversity of moral norms in different, far-off societies. Newton's theory of absolute space (the 'sensorium of God') had to give way once Einstein had published his special theory of relativity. The simple dualism of a unified mind and a single brain has had to be rethought in light of discoveries about the differing functions of the right and left cerebral hemispheres. The concept of an object's being at a determinate place at a particular time is challenged by the indeterminacy inherent in quantum mechanics. And the very categories themselves in terms of which we analyze language have had to be re-examined in light of the discoveries of cybernetics and the quantification of information. The discoveries and theories of science provide much of the driving force of modern philosophy. Indeed, it strikes me as a certainty that the greater part of modern philosophy simply would never have been conceived except as a response to the information explosion in the sciences. Dozens of philosophers have spent years writing volumes on the relativity of space and time, on multidimensional spaces, etc. only because of the stimulus of post-Newtonian physics. Countless philosophers today pursue issues in, for example, medical ethics, sociobiology, linguistics, and artificial intelligence only because of the *need* within the many sciences for help with conceptual puzzles.

Since the late twentieth century, science and philosophy have begun to forge new alliances. The historical parting of the ways is being reversed. The discipline of cognitive science, for example, is as much philosophy as it is psychology and linguistics. Researchers in artificial intelligence do not fit comfortably into the strict categories of engineer or philosopher, but often are a bit of both. Economists, too, particularly those doing research in welfare economics, straddle the historical boundaries between empirical research and ethics.[3]

One of the most pervasive, but mistaken, notions of our modern era, a notion which is a clear holdover from nineteenth-century views about knowledge, is that each of our beliefs may be assigned to one of

[3] In recent years, many new interdisciplinary philosophy journals have been launched, including *The Journal of Medicine and Philosophy* (1976), *Linguistics and Philosophy* (1977) [successor to *Foundations of Language*], *Law and Philosophy* (1982), *Economics and Philosophy* (1985), and *Biology and Philosophy* (1986).

three categories: either a belief is, or can be, judged true or false by scientific investigation, e.g. that aluminum has a lower melting point than tungsten; or a belief is a matter of ethics to be decided either by convention or the pronouncements of some religion or other, e.g. that it is wrong to institute affirmative-action policies in admitting students to universities; or, finally, a belief is essentially just a matter of opinion, e.g. that some particular work of art or piece of music or sculpture is better than some other one. The first of these categories is, often honorifically, called "objective"; the third, often pejoratively, called "subjective"; and the second, something neither quite the one nor the other, viz. "conventional" or "God-given".

Neither the exclusiveness nor the exhaustiveness of these three categories is given much credence by philosophers any longer. The lines of demarcation between these categories are not nearly as clear-cut as formerly imagined. More important, these categories certainly do not include all the ways we human beings have found to explore truth and to ground beliefs.

It is not my purpose here to explore matters of ethics or of aesthetics. It will suffice, for my purposes, simply to warn that pursuing the answers to ethical questions and to aesthetic questions is never *just* a matter of convention or opinion. Like science and metaphysics, these areas of perennial concern must be informed by empirical data and by human reason.

What is of particular concern to me is to challenge the mistaken idea that so many persons have of the potency of science to answer questions lying outside of ethics, religion, and aesthetics. Science can take us only so far. Whatever results issue from a scientific experiment will always fall short of answering all the questions we contrive for ourselves. Consider, for example, Descartes's theory of the relationship between minds and brains.

In the seventeenth century, René Descartes (1596-1650) gave expression to one of the most enduring theories of mind and brain. He argued that minds and brains are distinct – i.e. different kinds of – substances, in particular that brains are *physical* things, all of which take up (occupy) physical space, while minds are *mental* things and do *not* take up physical space. Each of us, when alive, is a curious amalgam, then, of two different *kinds* of things, or substances: a physical thing, i.e. a human body including its brain, and a nonphysical mind or psyche. The problem, then, became to try to offer a theory explaining the relationship between these two substances. How is it possible for these two, essentially different, kinds of things to interact in a causal manner? How, exactly, does a cut in your finger cause (bring it about that you feel) pain? Wounds, on this theory, are *physical* events, while pains, on this theory, are supposed to be *mental* events. How does something which occurs in space (an incision in your finger) cause something which is not in space (the pain you feel)? Or, taking another example, this time going the other way, how is it possible that your desire for a drink of water, your desire being – on this theory – something that exists in your mind and is thus not physical, causes your body, a physical entity, to rise up out of a chair, cross the room, turn on the kitchen tap, and lift a drinking glass to the stream of water issuing from the faucet? Descartes's successors, to this very day, struggle with these problems.

Other philosophers have offered quite different theories. Some have argued that Descartes's fundamental distinction, that is, between minds and brains, is misconceived. Some of these critics have argued, for example, that minds are not distinct from brains, that mental states and events are nothing other than states and events in the central nervous system. And needless to say, there are a variety of other theories, as well, which are variations on these two principal themes. Since the mid-twentieth century we have enjoyed (or suffered, depending on your attitude about these sorts of things) several different theories about the relationship between minds and brains.

The point of bringing up this embarrassment of riches is to focus on a certain naive attitude some persons bring to these debates. Every teacher who has ever introduced students to the problem of the relationship of mind and brain has learned to expect that some students will regard the question as one to be settled by empirical research in the laboratory. To these students, the question initially appears analogous to questions such as "What is the relationship between fever and infection?" and insofar as this latter question permits of empirical resolution, so, too, does it seem to them that the former, about the relationship between mind and brain, ought to permit of empirical resolution. These students express confident belief that – just given enough time, money, and resources – scientists will be able to decide such issues by their usual methods.

Nothing that emerges from the experimenters' laboratories can ever, or indeed could ever, settle the issue between dualists, persons who hold, like Descartes, that mind and brains are fundamentally different kinds of substances, and monists, persons who hold that there are not these two kinds of substances.

It has often been alleged that the reason science cannot provide a definitive answer to Descartes's puzzle is because science is essentially directed to exploring material, or physical, features of the world, that science is incapable of exploring nonphysical entities. But the explanation of the source of the difficulty is, I think, really much more profound. For it is simply false that science is essentially incapable of examining nonphysical things. Suppose, just for the sake of argument, that minds, or more exactly mental states and mental events, *are* nonphysical. Suppose, just for the sake of argument, that mental states and events are – let us say – weightless, have no exact position within our bodies, lack color, have no scents, etc., i.e. lack most, if not all, of the properties we usually find in physical entities. Nevertheless, even if all this were true, science could still explore and learn a very great deal about mental states and events. Science might learn, for example, whether it were possible for persons to 'think themselves out of' pain, whether bright patches of red in our visual fields are succeeded by green patches, whether certain odors might evoke certain memories, whether musical acuity correlates with mathematical ability, etc. As a matter of fact, these sorts of experimental researches are precisely the kinds psychologists regularly do pursue, and indeed do so without having settled the issue whether mental states and events are, or are not, brain states and events. In short, the inability of science to answer the question whether mental states are brain states does not arise because science is essentially incapable of examining nonphysical entities. For all we know mental states may be nonphysical entities. If they are, then science is, even now, examining nonphysical entities.

The real source of the difficulty in trying to decide the relationship between mind and brain is that in exploring the mind of a subject, psychologists have no non-question-begging test to tell them whether they are looking at some physical feature of that subject's central nervous system or at some other, nonphysical, feature of that subject which, although possibly correlated with some physical feature, may – for all that – be distinct from it.

Suppose a dualist and a monist were having an argument. Suppose the dualist were to offer the theory that pains *accompany* or *are caused by* certain kinds of nerve firings but are not literally those firings themselves, while her colleague, the monist, were to argue that the pains were not something *accompanying* the nerve firings but were, literally, the nerve firings themselves. They may perfectly well agree about the experimental data: that increasing the rate of firings increases the reported magnitude and duration of pain; that anesthetizing certain nerves blocks the pain; that bombarding the subject with intense so-called pink noise also blocks the pain; etc. What would any of this show about the correct answer to the question whether pains are physical events? It should be clear that none of this would settle the matter. The question, and its resolution, goes well beyond the experimental data.

This is not, of course, to deny that empirical data have a bearing on the issue. They do. But only up to a point. If there were no detectable *correlation* between pains and physical states, if pains seemed to occur randomly without any physical cause whatever, then we probably would be positively disinclined even to consider that they might be physical states. But we already know enough about the world to know that pains, more often than not, do occur when certain sorts of physical events occur. What experimental research tells us, then, is that there is not just a bare possibility of dualism's being false, but that there is empirical evidence which is consistent with another theory, viz. monism.

The empirical data which we have about minds and brains are consistent with (at least) two *different* theories of the relationship between minds and brains. And from an examination of the sorts of data which empirical research is capable of yielding, it is clear that no forthcoming or future data could ever settle the dispute. It is at this point that the problem, having originated within science, must go beyond science, to metaphysics.

How, if not by empirical research, is metaphysics to be pursued? We have already seen, in chapter 4, that Logical Positivists had been convinced that metaphysics is an impossibility and that any pretensions it might have to furnishing us knowledge are illusory. That particular opposition has pretty well, in time, damped down. But the challenge remains. We really must address the problem how, if not by empirical research, one can hope to answer the sorts of questions science must leave unanswered. To what might one take recourse?

I hope that the answer should already be beginning to become clear in light of what I have been saying in the previous three chapters about theories and underdeterminism. We pursue metaphysics by trying to construct the best theories we can to explain, i.e. to make sense of, the puzzling aspects of our world which science is incapable by itself of explaining. As we venture further from the empirical base, our theorizing becomes more difficult and less determined, or if you will permit me an ugly (but informative) phrase, our theorizing becomes *more underdetermined*.

Metaphysical theories about, for example, the ultimate relationship between minds and brains, theories which presuppose and try to accommodate all the empirical data of the experimental laboratory, but which also try to go beyond that data, are probably destined never to convince all rational persons. Being even more underdetermined than the scientific theories on which they ride piggyback, they probably always will be the object of dispute and of criticism. But this is simply our human condition. No matter how much empirical data we gather, however diligently and conscientiously, that data always will be insufficient to establish our scientific theories as true and always will be insufficient, to an even greater degree, to establish our metaphysical theories as true. Yet we human beings seem insatiably curious. Our knowledge has no natural stopping point. Most of us rebel at the very thought that our knowledge may be at an end. Even if the methods and tools of psychologists and physiologists cannot, in principle, tell us, for example, finally whether minds are, or are not, brains, we still persevere in wanting to know such a thing. And if science cannot tell us, then we will go beyond science, to metaphysics.

6.2 Vagueness

Metaphysical theories are proffered solutions to *conceptual* puzzles. What is the concept of *mind*? What is the concept of *person*? What is the concept of *identity*? What is the concept of *material object*? What are the concepts of *space* and of *time*? Etc.

It is not essential that we try to get very clear what a concept is. That exercise may be left for books on the philosophy of language and of mind. Let me say only this: persons have a concept of – let us take as an example – redness, if they are able, for the most part, to use correctly the word “redness” or some other word, in some other language, which means pretty much what “redness” does in English. To be sure, there are a fair number of things about which one could seek clarification in this brief exposition. But let me add only that, on this particular explication, concepts are not words themselves, but are expressible by words. Speakers of different languages, then, e.g. English and French, may have the same concepts although they will use different words to express them, e.g. “redness” and “rougeur”.[4]

[4] I have stated a quasi-sufficient condition for having a concept, but not a necessary one. Animals, e.g. dogs and cats, probably have certain concepts, but dogs and cats lack languages. Thus, having certain linguistic abilities

Bertrand Russell (1872-1970), in a famous paper, once wrote that "all language is vague" ([180], 84). He also wrote that "all knowledge is vague" (90), that words are vague (85), and "that every proposition that can be framed in practice has a certain degree of vagueness" (88). What all this comes down to is that the concepts which figure in the propositions we frame, i.e. in our beliefs, theories, musings, doubts, certainties, etc., are themselves vague. Vagueness in beliefs, in theories, etc. is traceable to vagueness in our concepts.

In what I am sure was intended as a facetious remark, Russell added, "whatever vagueness is to be found in my words must be attributed to our ancestors for not having been predominantly interested in logic" (84). Had he reflected upon this remark, I am sure Russell would have had to admit that history could not have been otherwise. For general terms, such as "red", "warm", "bald", and "tasty", to be useful, indeed to be learnable by persons not born already knowing a language, those terms *must* be somewhat vague. Concepts could not be taught, and could not be learned, except if they were vague.

A term, or a concept, is vague if there are particular cases (instances, things, etc.) which do not fall clearly inside or outside the range of applicability of the term. At what point exactly does an object cease being red? How much hair exactly must a man lose to be judged bald? In what manner and to what degree must a person hold theological beliefs and practice the dictates of some religion to be regarded as religious? Such questions admit of no precise answers and the concepts *redness*, *baldness*, and *religious* must be regarded as vague.

If most of our workaday concepts were not somewhat vague, we probably could never have learned them, and we certainly would have grave difficulty in applying them. Indeed we can easily see how utterly counterproductive it would be to attempt to reduce the vagueness inherent in the concept *bald*. Suppose we were to stipulate that anyone who had fewer than 2089 hairs on his scalp was to be regarded as bald. (This of course presupposes that we have already similarly stipulated where one's scalp leaves off and one's face begins. The difficulty thus begins to mushroom. What begins as a problem in making the concept *bald* more precise quickly also becomes a problem in revising the concepts of *face*, *scalp*, *hair*, etc. But we will not trouble

suffices in the case of human beings to show that those human beings have certain concepts, but those linguistic abilities should not be taken as the very same thing as having those concepts.

ourselves over this appreciable further difficulty, since we have problems enough with the very first level of reform.) On this new, refined, concept, some persons would clearly be seen to be bald, others clearly not so. But there will be cases for which we would have to count hair shafts to be sure. Now, given the sorts of taboos in this society, and given our sorts of interests, we are not much inclined to go about conducting an inventory of the precise number of hairs on an acquaintance's scalp. If we were to impose the imagined precision on the concept *bald*, we would, almost certainly, immediately supplement that precise, and now fairly useless, concept with a vaguer, more practical one, one which for all intents and purposes would simply duplicate our present concept of *baldness*. In short, any suggestion that we have vague words in our language, vague concepts in our conceptual scheme, only because our ancestors did not make the effort to be more precise would be wildly wrong. We have vagueness in our concepts because of the way we learn language and because some vagueness is positively required.

For a conceptual scheme to be workable it is essential not only that some concepts be relatively vague but that there also be mechanisms for reducing vagueness as the need arises. No legislator could possibly envisage all the ways there might be to create a public disturbance or to threaten the public health. At the turn of the twentieth century only a clairvoyant could have imagined the nuisance potential of portable stereo radios and the threat to health posed by the careless use of DDT or by the promiscuous behavior of an AIDS sufferer. Key terms in legislation – "public good", "intellectual property", "privacy", etc. – must always remain somewhat vague and thus open to later refinement in light of changed circumstances.

What is true of legalese is true as well of English at any stage of its gradual evolution. Our concepts are adjusted to allow us to cope with the world pretty much as we find it now. As our knowledge expands, as new technology appears, as our sense of right and wrong gradually changes over time, our concepts must be revised to allow us to operate with these changes. The *word* "justice" has not changed for a few hundred years, neither has the *word* "art", nor the *word* "death". But our *concepts* of justice, of art, and of death surely have changed in that period. It was not too long ago, for example, that a person whose heart had stopped beating was considered dead. Not any more. Nowadays a person whose heart has stopped beating may be considered, like persons who have temporarily stopped breathing, not dead but in a life-threatening situation which may warrant heroic resuscitative measures. In short, we have, in the last fifty or so years, in response to a changed medical technology, revised, not the word “death”, but the *concept* of death.[5]

6.3 Conceptual analysis

To solve a metaphysical puzzle is, ultimately, to make a suggestion, i.e. to offer a theory, about how we *do* in fact use a concept or – more often – how we *ought* to use (i.e. revise) a particular concept. When Hume, for example, began his examination of the endurance through time of material objects, he took care to insist that as we ordinarily conceive of a thing’s enduring we allow for the thing to change somewhat.[6] He asks us to imagine how we would respond to the situation of something’s changing counterfactually* (i.e. contrary to the way the world actually is) by the acquisition or loss over time of some small part. He anticipates that we would not, ordinarily, count that as destroying the thing’s identity, and takes that expected response as evidence of how we actually do use the concept. (We will return to this sort of technique, construing its counterfactual aspect as the describing of a possible world, in section 6.4. There we will examine Locke’s attempt to elicit our ordinary concept of *person*.)

This first kind of conceptual analysis, I will call “narrow analysis”. Some authors prefer the expression “pure analysis”. “Narrow analysis” simply means an assay of the standard or typical conditions for the applicability of a concept. Thus Hume’s preliminary remarks about endurance through time can be regarded as narrow analysis. And so, too, can the example Bradley and I offered in our book *Possible Worlds*, when we argued, in effect, that the concept of *knowledge* is a complex concept ([34], 183) and that it has the concept of *belief* as one of its constituents (23). Put another way, we argued that it is part of the *analysis* of the concept of *knowledge* that any case of *knowledge* is also a case of *belief*, e.g. to know that it is raining entails (among other things) believing that it is raining. The example we

[5] Certain English words have changed enough over a period of three hundred years so as to make reading some eighteenth-century philosophy problematic. One must recognize, for example, in reading Locke and Hume, that words such as “idea”, “impression”, “power”, and “necessity” simply do not mean to modern readers what they did to eighteenth-century readers.

[6] The passage from Hume ([101], 255-6) is quoted below, p. 331.

chose is perhaps more controversial than one would like for illustrative purposes. Nowadays I think I would be more inclined to choose as an example the concept of *triangle*, and would argue that it is part of the *analysis* of the concept of *triangle* that all triangles have three sides.[7] Such knowledge as this is standardly called "analytic knowledge", and on the account preferred by some philosophers at least, Bradley and I certainly among them, would be deemed to be a priori knowledge.[8]

But there is a second kind of analysis, one which may be called "broad analysis", which goes far beyond the limited descriptive nature of narrow, or pure, analysis. In this latter kind of analysis there is a significant component of revision. (Recall the quotation from Strawson in chapter 2, p. 23.)

When we read Hume's discussion of the concept of *causality*, and more especially Kant's response to that discussion, we detect a far greater degree of revision than appeared in the discussion of endurance. Kant, in trying to figure out how and why we make causal attributions, given Hume's claim (which Kant accepted) that causal con-

[7] There are, of course, additional concepts involved as well, viz. that triangles have straight sides, and that the figure is closed. The concept *three-sided* is, then, but one constituent of the concept *triangularity*.

[8] Let me quote from my explication of a priori knowledge which appears in the Glossary at the back of this book: "When philosophers say that a statement can be known without experience, they mean that no particular experience of the world, *save perhaps learning a language*, is necessary to be able to figure out the truth or falsity of the proposition." The italicized qualification is essential. Learning a language can only be through experience. We are not born knowing any language. But learning a language, learning under what conditions it is appropriate to apply words such as "red", "bald", "triangle", "knowledge", etc., is not to have empirical knowledge of this world. Learning English is not, for example, sufficient to inform us whether all triangles are red, or whether all squares are fragile, etc. I might learn, for example, to speak Swahili, and might learn the Swahili words for certain foods I never have laid eyes upon. But merely learning to use these words would certainly not put me in a position to know whether these various foods are ever served at the same meal together, or whether there might be religious strictures barring their appearing on the same table together. To come to know the latter, one would have to go beyond mere knowledge of the meaning of (Swahili) words, to *empirical* knowledge of the mores of the Swahili people.

nections can never literally be perceived – i.e. that all that is perceivable is *sequences* of events – proposed a new concept of *causality*: one which, like his analysis of space and time, theorized that *causality* is imposed on the data of sense 'by the mind'.[9]

When one undertakes to revise a concept, as Kant has done in the case of *causality*, the task is considerably more challenging, and likely to draw fire, than when one tries merely to report how we typically use a concept. Contributing to the difficulty is the fact that there is no one way to go about inventing and arguing for changes in our conceptual scheme. Nor are there probably just a few ways. There are many different ways, some of which are historically and currently stylish, others – no doubt – not yet even imagined, i.e. others are surely to be invented, developed, polished, and pursued by our descendants long after all of us are dead.

There is as much, or as little, method in the practice of metaphysics as there is in science. The commonly broadcast claim that there is something called 'the scientific method' is for the most part – as I have tried to show earlier – more fable than fact. Whatever scientific method exists is not something to be captured in a set of recipes, either by Bacon or by Mill or by a modern successor. What method there is to the practicing of science is something learned by apprenticeship, by imitation, by trial and error, by imagination, and by exposing one's scientific work to the scrutiny and criticism of fellow scientists. The practice of metaphysics proceeds in a parallel fashion.

The practice of metaphysics must be learned by apprenticeship, by reading the writings of metaphysicians, by attending to criticisms of those writings, and by daring to construct theories on one's own, theories which – just on the basis of probabilities – like most new theories, will prove to be, in one way or another, defective. Metaphysics, like science, is not for the fainthearted, and if it is to be done well, is certainly not for the dogmatic.

But if there is no particular method to be followed in metaphysics in our attempts to *generate* revisions to our concepts, there are, nonetheless, certain desiderata to be looked for in *judging* whether a proffered revision of a concept is to be accepted or rejected. Rudolf Carnap (1891-1970) has provided us with one of the best, most insightful discussions of the features by which to judge the worthiness of a philosophical reconstruction ([45], 3-8). Carnap uses the term "explication"

[9] See the quotation on p. 13 and the discussion on p. 30.

to describe the process of revising a concept, what has here been called “broad analysis”. The concept to be revised, he calls the “explicandum”; and the new concept, the one which *replaces* the original, he calls the “explicatum”. The latter, the new concept, the explicatum, is supposed to be less vague than the original concept, the explicandum. But there is no question of the explicatum’s being perfectly precise. The explicatum is devised to be an improvement; it is not to be thought of as a finished product suitable for use in all subsequent circumstances.

Other authors use different technical vocabulary. Where Carnap uses the term “explicandum”, other authors sometimes use another Latinate term, viz. “analysandum”, while still others prefer the technical English phrase “pre-analytic concept”. And where Carnap speaks of the “explicatum”, other authors speak of the “explicans”, the “analysans”, the “reconstruction”, the “explication”, and the “analysis”, in the last two instances using the terms “explication” and “analysis” both for the process of philosophical reconstruction and for the product of that reconstruction.

Since the 1950s, the terms “analysis” and “analytic philosophy” have achieved a remarkable philosophical vogue. But that fashionableness has been accompanied by an unfortunate ambiguity. Occasionally “analysis”, without any accompanying adjective, is used as I have used the expression “narrow analysis”; very often, however, it is used as I have used the expression “broad analysis”, or “revisionary analysis”, or as Carnap has used the term “explication”. On balance, when other philosophers use the term “analysis” *tout court*, they probably mean the latter sort of analysis, viz. “explication”. Although the truncated expression “analysis” is not entirely apt, and perhaps even a little misleading, it is so well established within the philosophical lexicon that it would be futile to try to avoid it. So hereinafter, “analysis” will be understood to mean “explication”.

Carnap argues that in judging a philosophical analysis (i.e. an explication or reconstruction), there are four factors to be considered, viz., that the explicatum (the revised concept) is to be

1. *similar* to the explicandum (the pre-analytic concept)
2. as *exact* as possible
3. *fruitful*
4. as *simple* as possible.

There can be little question that in assessing and criticizing philosophical analyses (reconstructions), these four factors do play a pivotal role. Critics will praise or fault an analysis according to the degree they judge the analysis to satisfy these requirements. But there is no mechanical or precise way to go about quantifying these requirements, and disputes among philosophers can be fierce as to the relative merits of an analysis. Thus, for example, as we saw earlier, the formalists admire and promote a particular theory of explanation, the so-called covering-law model, which reconstructs explanations as arguments, in which universal or statistical laws, along with statements of antecedent conditions, figure as premises, and the statement describing the event to be explained figures as conclusion (e.g. recall Hempel's example [p. 36] of a car radiator cracking).[10] This analysis is prized by the formalists for its exactness (Carnap's second requirement) and its relative simplicity (Carnap's fourth requirement). But other philosophers are adamantly opposed to the covering-law model, arguing that it does violence to our pre-analytic notion of explanation which is heavily context-dependent and turns on the background knowledge of the person seeking the explanation and on whether that person succeeds in understanding the explanation. In short, critics of the covering-law model protest that the explicatum departs too far from the explicandum, i.e. that the model fails on the first of Carnap's requirements.

Philosophical disputes, such as that between the covering-law theorists and their opponents, are inevitable. The trouble is that the various desiderata in a philosophical explication are usually in conflict with one another. Exactness and simplicity are often purchased at the price of severely constricting our pre-analytic concept. Such clashes are, in the main, inescapable. Our pre-analytic concepts often get into trouble as they are extended to handle cases beyond the 'typical' ones, i.e. they are not suitably clear or precise enough to comprehend new, problematic cases. But there never is, nor could there be, any one way to modify a concept, and there is bound to be disagreement about the benefit of suggested changes.

[10] Beware not to confuse *explanation* with *explication*. We explain, for example, historical events, biological processes, the means to start a car, and – on occasion – we explain the uses of a term (as I am doing right here, now). But we never explicate historical events, biological processes, etc. In reconstructing a concept, in offering a theory as to how it might be revised, we explicate that concept, we do not explain it. Thus, it is possible, as we see here, to explicate the *concept* of explanation. To do so is not to explain explanation.

All of this presupposes that concepts are not the sorts of things which are 'out there' in some sort of absolute realm not of human making. Concepts are not the sorts of things which we human beings discover and either, as it were, comprehend 'correctly' or – because we make a mistake, through carelessness, inattention, confusion, etc. – comprehend 'incorrectly'. That sort of theory of concepts has been so thoroughly dismissed in modern philosophy that it no longer bears refutation. The concepts human beings use are not fixed entities to be mined in an intellectual realm, but are human inventions, although of course – as mentioned earlier – there may be physical constraints placed on what sorts of concepts we are capable of inventing. (As I said before, it is possible that human beings are so physically constructed, 'hard-wired' in our central nervous systems, as to be capable of forming and entertaining only certain sorts of concepts. Animals – e.g. Thomas Nagel's example of bats – may be incapable of forming the concepts we human beings use; and we, in turn, may be quite incapable of forming the concepts animals use.) In *inventing* and *revising* concepts, there can be no question of our 'getting it right' in the sense of our reconstruction being *true*. Philosophical analyses, although having elements which may be judged true or false, are not, overall, to be regarded as the sorts of things which are true or false. Whether an explicatum is similar to an explicandum can, to a certain degree, be considered something admitting of a judgment which is true or false; similarly the judgment whether an explicatum is fruitful or simple or exact may, too, be regarded as something having aspects of truth and falsity. But to judge an explication as fruitful, or simple, or exact, does not address the questions whether the concept is fruitful *enough*, whether it is *simpler* than rivals, and whether its exactness is *purchased* at the price of its simplicity and similarity to the pre-analytic concept. For these latter sorts of judgments, one cannot argue simply on the basis of 'fact' or pretend that one's claim is somehow manifestly true or that a competing explication is manifestly false. In making these latter judgments which are inherently part of judging the worth of a philosophical analysis, many factors *other* than truth and falsity come into play.

(It bears remarking that there is one particular term which is often also invoked in these contexts, but which may prove a pitfall for the unwary, viz. "intuition". Unless one is aware that philosophers use this term in a specialized, technical sense, confusion is bound to result. When philosophers speak of "pre-analytic intuitions" and "pre-philosophical intuitions", and say of an analysis that it is, or is not [as the case may be], "counterintuitive", they are *not* using the term "intuition" as it is ordinarily used, to mean something like "knowledge prior to, or independent of, experience", as for example, when biologists might say that a bird's ability to build a nest without having been taught to do so is intuitive [or instinctual]. In the context of philosophical analysis, or reconstruction, when philosophers talk of intuitions, they refer to our judgments about the pre-analytic concept, the *analysandum*. And when they say such things as "the analysis is counterintuitive", they mean that the proffered reconstruction strongly departs from the original concept. Thus, for example, were a philosopher to complain that some particular analysis of mind was counterintuitive, she would not be objecting that the analysis contradicted some in-built, or a priori, knowledge, but would – rather – be making the more reasonable claim that the analysis was very unlike the original, pre-analytic, concept.[11])

Philosophical analyses of concepts are nothing like dictionary definitions. Dictionary definitions are esteemed for their brevity. Philosophical analyses are anything but brief. Carnap's discussion of the requirements for a philosophical explication occurs in his book *Logical Foundations of Probability* (1950) in which he explicitly says he is trying to explicate one particular concept of probability. The ensuing explication of what he calls "logical probability" consumes well over 500 pages. Or, again, when Gilbert Ryle endeavored to offer his analysis of mind in *The Concept of Mind* (1949) his explication ran to more than 300 pages. And John Rawls's analysis of justice, *Theory of Justice* (1971), spans 607 pages. To revise a concept is no easy or trivial matter. One must try to get clear how the concept has been used, how it runs into trouble for certain cases, and how altering it in a cer-

[11] The strongest case I know of a philosopher arguing for the importance of intuitions in judging a philosophical analysis occurs in Saul Kripke's *Naming and Necessity*: "… some philosophers think that something's having intuitive content is very inconclusive evidence in favor of it. I think it is very heavy evidence in favor of anything, myself. I really don't know, in a way, what more conclusive evidence one can have about anything, ultimately speaking" ([116], 42). But even with this said, Kripke proceeds to argue that intuitions are not inviolable. For he immediately states that persons who find a particular philosophical thesis – that there are accidental* properties – unintuitive have their intuitions "reversed". Clearly, Kripke thinks both that some intuitions are to be preferred to others and that some intuitions can be successfully challenged by cogent argument.

tain fashion might be thought to provide a way out of the difficulties.

If philosophical analyses are not like dictionary definitions, they are not like proofs in mathematics either. Unlike what I have called "narrow analyses", philosophical analyses (explications) are not the sorts of things whose correctness (validity) may be demonstrated a priori. There is far more empirical content in philosophical analyses than is usually recognized. Because the two different kinds of analysis have not always been clearly distinguished and because few philosophers ever conduct original scientific research, e.g. because few philosophers ever themselves mix chemicals, construct electronic equipment, peer through microscopes or telescopes, dissect animals, unearth ancient pottery, or drill boreholes, it has been easy to form the mistaken idea that philosophy is an a priori science. And probably more than a few philosophers themselves have believed just this. Some persons, in noting that philosophers typically conduct no empirical research, have come to believe that philosophy is just simply the product of deep thought, that philosophy – at its best – springs from pure, unaided reason and that the less contaminated it is by crass facts, the more it aspires to independence of 'the facts', the better it is. The truth is, however, that any such philosophy would be grievously impoverished. If philosophers themselves do not conduct empirical research it is only because they depend, secondhand, on the empirical research of others to infuse their own theories. We have here something of a separation of labors, but nothing like an exclusivity of objectives or interests.

Metaphysical analyses of personal identity, for example, depend heavily and crucially upon certain *empirical* facts about material objects (e.g. their impenetrability, their endurance through time) and upon certain empirical facts about memories (e.g. that memories are causally related to witnessed events in one's own lifetime). Similarly, philosophical analyses of *art*, of *labor*, of *agency*, of *free will*, of *miracle*, and of *justice*, etc., all presuppose an enormous background of *empirical* facts. One cannot get far in discussions of justice, for example, without a host of empirical presuppositions, both psychological and sociological: e.g. that there is a scarcity of material goods; that human abilities, desires, and opportunities are not distributed equally; that certain desires conflict with those of other persons; that disease and physical handicaps afflict some, but not all, persons; that individual lifetimes differ greatly; that knowledge and information are commodities which are not universally accessible; and that decisions are often taken in ignorance of what other persons are doing. Thus, in trying to devise a philosophical explication of justice, no philosopher ever could hope to fashion such a theory a priori or hope to fashion a theory which would be applicable in every conceivable set of circumstances. Quite the contrary, what is involved in fashioning a philosophical theory is the desire that that theory should be applicable to this particular world. Metaphysical theories must have substantial empirical content; they must, that is, be tailored to this world, not to any and every possible world.[12]

And yet, paradoxically, one of the most powerful tools philosophers sometimes use in creating and testing their theories is to put the concepts at play under stress by asking what we would want to say if the circumstances were markedly different from the circumstances that ordinarily prevail. This technique is so surprising, and yet so useful and widely practiced, that it demands particular scrutiny.

6.4 Possible worlds

The aim in developing a philosophical theory is to clarify and possibly revise some of our concepts. How might we do this? One way, particularly favored among metaphysicians, is to place the concept under stress, to subject it to a kind of test wherein we ask whether or not we would want to persevere in applying the concept to *counterfactual* cases, cases which are sometimes far from ordinary; indeed, in many instances, which are physically impossible.

Thus, for example, in one of the most famous passages in philosophical literature, John Locke (1632-1704), in trying to analyze the concepts of *person* and *personal identity* writes:

> … should the soul of a prince, carrying with it the consciousness of the prince's past life, enter and inform the body of a cobbler as soon as deserted by his own soul, everyone sees he would be the same person with the prince, accountable only for

[12] "… many concepts of philosophically central interest are collage-like: they are internally diversified *combinations* of logically separable elements that are held together by the glue of a theoretical view of the empirical facts. Such concepts rest in an essential way on an empirically-based, fact-laden vision of how things work in the world. … Our concepts are not framed to suit *every possible* world but in significant measure adjusted to *this* one" (Rescher [169], 120).

> the prince's actions; but who would say it was the same man? The body too goes to the making the man and would, I guess, to everybody, determine the man in this case, wherein the soul, with all its princely thoughts about it, would not make another man: but he would be [taken to be] the same cobbler to everyone besides [i.e. except] himself. ([124], book II, chap. XXVII, §15)

Unfortunately, Locke's prose can be maddeningly obscure at times. This paraphrase of the passage may help to make it clear what (I think) Locke is trying to say:

> Suppose the consciousness of a deceased prince were to enter and infuse the body of a cobbler immediately upon the cobbler's own consciousness leaving his body. Anyone who knew of this transference would immediately regard the living person not as the cobbler, but as the prince. But were someone ignorant of the transference, then, in judging from the evidence of the physical body, he would take the person to be the cobbler.

All of this is more than a bit strange. So far as we know, consciousnesses, souls, psyches (call them what you will) do not flit from body to body. Indeed, so far as we know, not only do such things not happen, the transference of consciousnesses is physically impossible. Why, then, should Locke, in trying to understand the concepts of *person* and *personal identity*, even consider such an outlandish counterfactual scenario? The explanation is immediately forthcoming:

> I know that in the ordinary way of speaking, the same person and the same man stand for one and the same thing. … But yet when we will inquire what makes the same *spirit*, *man*, or *person*, we must fix the *ideas* of *spirit*, *man*, or *person* in our minds; and having resolved with ourselves what we mean by them, it will not be hard to determine in either of them or the like when it is the *same* and when not. ([124], book II, chap. XXVII, §15)

Locke is here using the term "man" much as we would today use the term "human being", i.e. as designating a certain kind of physical creature, a member of the species *Homo sapiens*. Locke notes that the expression "same man" (or "same human being") and the expression "same person" generally – or as he puts it, "in the ordinary way of speaking" – stand for one and the same thing. Of course he is right: whenever you judge some man (human being) to be the same man as, for example, when you judge the clerk in the drugstore to be the same *man* as the man who used to work in the bakery, you could equally well say that this *person* is the same person who used to work in the bakery. In the ordinary way of speaking, "same human being" and "same person" may, stylistic considerations aside, be considered *interchangeable*. You could even say that this *man* who is now working in the drugstore is the same *person* who used to work in the bakery. But in spite of this – and this is the crux of Locke's point and of his strange example – the two concepts are not after all the same. The concept *human being* – Locke tries to show through his counterfactual supposition – is different from the concept of *person*. Although all human beings may, as a matter of fact, be persons and all persons may, as a matter of fact, be human beings, nonetheless the concepts of *human being* and *person* are different.

Suppose, just for the sake of an example, that every creature which has kidneys also has a heart and, conversely, that every creature with a heart has kidneys (this example is Quine's [162], 21). Suppose, further, that you wanted to argue that the concept of *having kidneys* is not the same concept as *having a heart*. One thing you could not do would be to display some creature which has the one kind of organ but which lacks the other, for, by hypothesis, there are no such creatures. To demonstrate the conceptual difference between *having kidneys* and *having a heart*, you might take recourse to imaginary (counterfactual) cases. You could say something of the sort: "Imagine an animal which has blood and an organ to pump that blood, but in which there is no specific organ which filters waste products from the blood. Instead, waste products pass directly through the walls of the blood vessels, through the surrounding flesh, and are evaporated on the surface of the skin." The described animal is, of course, a creation of science fiction, i.e. does not exist so far as we know. But whether it exists or not, it serves your purposes admirably. The mere *logical possibility* of its existence suffices to show us that the concept *having kidneys* and the concept *having a heart* are not the same concept.

Now, of course, it must be admitted that we knew this all along. None of us for a moment was tempted to confuse, or conflate, the two concepts of *having kidneys* and *having a heart*. We knew these concepts were distinct before, and indeed without, your telling the science-fiction tale we just related.

But how, exactly, did we know this? We knew it because, even without perhaps ever having encountered a creature which had the one kind of organ and not the other, we were easily able to imagine such a creature. We could imagine, that is, that one of the two concepts might be applicable to some one creature and not the other one as well. That trivial piece of imagining suffices to demonstrate that the two concepts are distinct.

Now we can see what Locke was up to. Perhaps every human being we have ever encountered has been a person; perhaps every person we have ever encountered has been a human being. But are the concepts *human being* and *person* the same concept or not? (The fact that they are expressed by different English words is irrelevant to making the decision. The two words "asteroid" and "planetoid" are certainly different: the former has eight letters, the latter nine; etc. Even so, differences of expression aside, the concept of *asteroid* is the very same concept as *planetoid*.) Locke tries to show that there are certain describable counterfactual cases which, if they were to obtain, would be ones to which one of the two concepts, *human being* and *person,* would apply and the other one not.

If the mind of the prince were to enter the body of the cobbler, we would have such a case. For if this were to happen, and we were to know that it happened, we would – Locke confidently predicts – know that although this is now the body of the cobbler, i.e. this *human being* is the human being who used to be the cobbler, this *person* is the person who used to be the prince. (You may find yourself disputing Locke's prediction about how you would interpret such a case. You may, for example, be disinclined to conceive of minds – or personalities – in such a way that they could even be the sorts of things which might migrate from one body to another. We will return to your misgiving later. For the moment, let us confine ourselves to trying to understand what is thought to be shown by such counterfactual examples.)

Contemporary philosophers have borrowed a piece of terminology which was popularized in the seventeenth century by Leibniz, but which he, in turn, had adopted from Scholastic (medieval) philosophy.[13] The scenes and situations depicted in these short counterfactual

[13] Leibniz's phraseology is familiar, even in popular culture: "Why did God create this world? Because this is the best of all possible worlds". His actual words were: "There were an infinity of possible ways of creating the

(or science-fiction) tales, which are used to place selected concepts under stress, have come to be called by many contemporary writers "possible worlds".[14] When Locke considered the *possibility* of the prince's personality migrating from the prince's body to that of the cobbler's, he was not describing this world (i.e. the actual universe) but an imaginary world, a possible world different from the actual one.

Nowadays, the technique of probing concepts by subjecting them to stress within a (described) possible world is commonplace. Contemporary philosophical literature abounds with such examples. Friedrich Waismann, in an attempt to show that the concept of *physical object* does not entail the concept of *physical impenetrability*, asks us to consider a possible world in which two chairs occupy the same place at the same time ([209], 201-2). John King-Farlow, in an attempt to show (among other things) that speaking a language does not require the speaker to have a concept of *self*, asks us to consider a possible world in which trees describe the scenery by rustling their leaves ([110]). Sydney Shoemaker, in an attempt to show that time can pass with nothing whatsoever happening, asks us to consider a possible world in which specific regions are periodically subject to total freezing (i.e. subject to total so-called suspended animation) ([191]). Anthony Quinton, in an attempt to show that not all physical objects need stand in spatial relations to one another, asks us to consider a possible world in which any number of persons, upon falling asleep in England, share a dream of being on a tropical island ([164]). And Peter Strawson, in an attempt to show that certain non-tangible objects can be re-identified over time, even though not being continuously observed, asks us to consider a possible world consisting solely of sounds ([200], chap. 2). (We shall return to some of these examples in subsequent chapters.)

world, according to the different designs which God might form, and each possible world depends upon certain principle designs or ends of God proper to itself" ([120], 36). And, "It follows from the supreme perfection of God, that in creating the universe he has chosen the best possible plan … The actual world … must be the most perfect possible" ([121], 325).

[14] Some authors use other vocabulary. For example, Swinburne ([202], 167), uses the term "myth" instead; King-Farlow talks of "parables"; and still others talk of "fables". But many, of course, persist with "counterfactual situations".

6.5 Methodological concerns

Like every other currently practiced philosophical method, that of examining concepts by subjecting them to stress within a possible-worlds story is an object of controversy. There are two main concerns.

The first has to do with the *limits* of the method. The technique of putting a concept under stress in a possible-worlds setting may be used modestly – as a tool for discovering how we actually use that concept – or aggressively – as a means to promote a particular revision of that concept. Nicholas Rescher is relatively sanguine about using the method in the first way, i.e. as a tool of narrow analysis, but has severe misgivings about using it in the latter revisionist manner.

> … the analyst must take care not to press his would-be clarification beyond the cohesive force of the factual considerations that hold together the concept as such. And this has definite implications regarding the usability of the science-fiction type of thinking [i.e. a possible-worlds story]. … A science-fiction hypothesis can effectively bring to light the significant fact *that* certain of our concepts are indeed multi-criterial [depend on several logically independent factors] and rest on empirical presuppositions. But what this method cannot do is to serve as a basis for precisifying [i.e. making more precise] our *existing* concepts, because the supposedly more precise account that results in these circumstances will not and in the nature of the case cannot any longer qualify as a version of the concept with which we began. ([169], 113-14, 115)

Where Carnap, we saw earlier, was enthusiastic about making a concept more precise, Rescher, in contrast, is considerably more hesitant, particularly when that revision comes about through the telling of a possible-worlds tale. Rescher worries that the kind of revision that often results in this latter instance is especially prone to sever just those essential empirical roots which give our actual concepts their usefulness. His uneasiness is well-founded. But his critique does not end in a wholesale rejection of the method. He cautions: “The point of these observations is not to advocate an unbudging conservatism in the conceptual area. No doubt there might conceivably be substantial advantages to giving up some of our concepts in favor of others” ([169], 115-16).

Whether the result of some proposed philosophical reconstruction departs unacceptably far from our original concept is something that can be known only on a case-by-case basis; it can never be determined in advance, and certainly not solely on the basis of the method used to generate that proposed revision. Examining and revising our concepts by telling possible-worlds stories has its risks, and those risks unquestionably increase as one passes from merely assaying a concept to replacing it. But carrying a risk does not render a philosophical technique useless. It merely entails that its results must be judged individually and never accepted or rejected in a blanket fashion.

The second concern over the use of possible-worlds stories in the practice of philosophy stems from a heightened examination, and more than just a little criticism, of the concept of *possible world* itself. There have been in recent years a great many books (see e.g. [126], [122], and [72]), and a much greater number of journal articles, devoted to such questions as: “Just what *is* a possible world?” “What are the contents of a possible world?” “In what sense might a possible world be thought to exist?” “Are possible worlds logically prior to propositions, or are propositions logically prior to possible worlds?”

We need not, however, have settled views about such esoteric subtleties in order to utilize the technique of telling possible-worlds tales. One need not settle all philosophical issues, or even for that matter express an opinion about some, to proceed with others. This is true even when the concepts invoked are fundamental and at the core of one’s method.[15]

Otto Neurath (1882-1945), in the midst of (what I regard as) a misguided argument against metaphysics, did offer an insightful, now famous analogy:

> What is first given us is our ordinary language with a multitude of imprecise, unanalysed terms. … We are like sailors who have to rebuild their ship on the open sea, without ever being able to dismantle it in dry-dock and reconstruct it from the best components. … Imprecise ‘verbal clusters’ are somehow always part of the ship. If imprecision is diminished at one place, it may well reappear at another place to a stronger degree. ([143], 91-2)

[15] Recall that one does not need a rigorous concept of, e.g., *number* in order to be able to do quite sophisticated work in arithmetic and other branches of mathematics.

The sailors cannot begin with an unsheathed keel (at the foundation as it were) and build anew. All they can do is gingerly replace parts from time to time, taking care not to sink the ship as they proceed.

In doing philosophy, we must start with a great many assumptions, techniques, and with a history. Of course any and every bit of this is eligible for examination and eventual revision, or even rejection or replacement. But what we cannot do is to begin with nothing and build 'from the bottom up'.

It must be conceded that the concept of *possible world* needs elucidation (explication), that there are many, as yet, unsolved puzzles about the concept. No one could read current journal articles and be in the slightest doubt about this. But none of this shows that the concept cannot be used with great success in the explicating of selected *other* concepts. One does not have to have a sophisticated theory of possible worlds to invoke the concept of *possible world* and to get much mileage out of that concept in attempts, for example, to probe the concepts of *time*, of *physical object*, of *cause*, etc. Indeed, legions of philosophers have not felt themselves in the slightest deterred by certain unclarities in the concept of *possible world* from using it in just these sorts of pursuits. Although some considerably more refined concept of *possible world* may be necessary for technical advances in such fields as logic and the philosophy of language, an *intuitive* or ready-at-hand concept of *possible world* has been and remains adequate for use as a tool in explicating many metaphysical problems.

For example, the concept of *possible world* needs no apology when it is used in explicating the concept of *person*. A great number of philosophers, beginning with Locke and continuing through the present day, have implicitly or explicitly invoked the concept of *possible world* for this very purpose. But their doing so then raises the question why anyone should want, or for that matter should even be tempted, to refine the concept of *person* along the lines we have seen emerging in Locke's discussion. Why, if – as earlier claimed – the two concepts *human being* and *person* may always, or nearly always, be interchanged, might we want to try to distinguish them?

To this last question, there is no one answer, nor are any of several possible answers absolutely straightforward. Again, we find ourselves examining the very reasons some of us are attracted to metaphysics and others not.

The concepts, the beliefs, the theories and the myths each of us operates with are to a large extent inherited from generations of ancestors and from the culture in which we live. There is no compelling *necessity* that any of us ever should distinguish between *human beings* and *persons*. It is easy to conceive of a society (note here how natural it is to fall into yet another, brief possible-worlds tale) in which there were no such distinction and, moreover, in which it never occurred to anyone to make such a distinction or even to ask whether it might be useful to make any such distinction. We can imagine a society in which all human beings were regarded as persons, all persons were regarded as human beings, and that was the end of it. But the simple fact is, that is not this society. (The truth of the matter is that *this* world is not such a world, i.e. this world is not the *possible* world just described.) For in this society, there is a long history of philosophers, lawyers, scientists, novelists, etc. who have asked, and will continue to ask, "What is a person?" and who have not assumed that the answer is a foregone conclusion: "a human being". Our philosophical, historical, mythological, and literary writings are filled with such examples: in Plato's *Phaedo* where – in trying to solve the problem of knowledge – he hypothesized that persons exist prior to birth; in some religious writings, where persons are claimed to be able to survive bodily death; and in our myths, e.g. where Merlin was supposed to have been able to inhabit the bodies of animals. In his philosophical novel *You Shall Know Them*, Vercors has his protagonist, Douglas Templeton, intentionally kill his own nonhuman offspring, the product of his mating with an apelike creature. The question posed by such a (possible-worlds) fable is: "Has Templeton killed a person?" If he has, then he is guilty of murder; if not, then he has merely killed an animal.

Anyone who can be engaged by the play of concepts in Vercors's novel has implicitly already made the distinction between being a *human being* on the one hand and a *person* on the other; has recognized that it is conceptually possible for these two concepts not to apply to the same creature; and has allowed himself to be receptive to a discussion of what, exactly, we are to understand by the concept *person*. The eventual answer is not to be found by a kind of philosophical excavating, i.e. the answer is not 'out there' for the discovering. The answer, if it is to be forthcoming at all, will be the product of suggesting ways we might want to extend our intuitive (or pre-analytic) concept. We may even want to try to imagine how we might react if placed in Vercors's imaginary, possible, world. Confronted with the offspring of a human being and an apelike creature, would we be inclined to regard that offspring as being more like a person or more like an animal? What sorts of factors would enter our decision? Perhaps appearance: how much does it look like a human being?; perhaps intelligence: how clever is this creature?; perhaps linguistic abilities: can it learn a language?; etc. But this is hardly the end of it. Other authors have taken more extreme cases and press the concept of *person* even further. In Arthur C. Clarke's *2001*, Hal (*H*euristically programmed *AL*gorithmic computer) seems to have 'a mind of its own', so much so that it turns rogue, subverts commands, mutinies, and murders human beings, and when its crimes are uncovered, pleads not to be disassembled and despairs at the prospect of its impending loss of consciousness:

> "Dave," said Hal, "I don't understand why you're doing this to me. . . .I have the greatest enthusiasm for the mission. . . .You are destroying my mind. . . .Don't you understand? . . .I will become childish. . . .I will become nothing. . . ." ([48], 156)

Hal is made of the typical stuff of computers: wires, transistors, etc. Hal, clearly, is not a *human being*. But might Hal be a *person*? One of the principal goals of contemporary research in artificial intelligence is to come to learn the operations of human cognitive processes well enough so that they can be replicated in a machine. If that comes to pass, what is now regarded as just a possible world – a computer threatening and pleading with us – we might discover is really a future period of this, the actual world.