Part I

Extended Essays

Causation

Making something happen, allowing or enabling something to happen, or preventing something from happening. Mental and extra-mental occurrences, of all spatial and temporal dimensions, great and small, have causes and are causes. Our awareness of the world and our action within the world depends at every stage on causal processes. Although not all explanations are causal, anything that can be explained in any way can be explained causally. Like other metaphysical concepts, the concept of causation applies very broadly. Yet this fundamental concept continues to elude metaphysical understanding. While there is some general philosophic agreement about causation, there is also considerable disagreement. Causal theories of knowledge, perception, memory, the mind, action, inference, meaning, reference, time, and identity through time, take a notion as fundamental that philosophers understand only incompletely.

HUME is the dominant philosopher of cause and effect. A running commentary on Hume's views and arguments, pro and con, could cover most contemporary philosophical concerns with causation (Hume, 1739, esp. Bk. I, Pt. III; Hume, 1748, esp. sects. IV, V, VII). According to Hume, it is not the experience of an individual causal transaction, but experience of other transactions, relevantly similar, that provides what causation involves in addition to priority and contiguity. Experiences of regularities or constant conjunctions condition our expectations. We project our conditioned feelings of inevitability on external objects as a kind of necessity that resides in the objects themselves (*see* Hume, 1748, sect. VII). Limitations of space preclude extensive quotation and discussion of these and other primary texts.

A number of paragraphs in this entry begin with the statement of a view about causation. The next sentence then classifies the view as *prevailing, majority, controversial*, or *minority*. Some of these classifications may themselves be controversial. Their purpose is only to help organize the entry.

Continuous causal paths connect causes with their effects. This is a prevailing view. Causes and effects are often not contiguous. A switch on the wall is distant from the electric light overhead that it controls. Pulling a button on an alarm clock makes it ring six hours later. The New York performance of three musicians in 1937 contributes causally to what one hears on the Perth radio in 2007. Although intervals of space, time, or space–time separate the causes and effects in these examples, spatio-temporally continuous causal paths connect them. The path has no spatial or temporal gaps or breaks. (A rigorous definition of CONTINUITY requires the notion of a *limit* found in calculus textbooks.) The path is causal because for any two positions, *a* and *c*, on the path, there is an intermediate position *b* on the path such that either something at *a* causes something at *b* that causes something at *c*, or the causation runs in the other direction, *cba.* An explanation of what constitutes a causal path that does not use the notion of causation would serve as a reductive definition of causation. The explanation above, which uses the notion of causation explicitly, serves only to state a spatio-temporal necessary condition of causation.

Causes and effects are events. This is a majority view (*see* DAVIDSON, 1980). Idiomatic speech often mentions something other than a change, or non-change, or occurrence, as a cause or effect, as in "Richard makes me furious." The question is whether an available paraphrase such as "Reading what Richard writes makes me become furious" brings events back into the picture as causes and effects. If both causes and effects are always of the same kind, then causal paths can continue indefinitely both from the past and into the future. On the other hand, the strategy of reducing all causal statements by paraphrase to statements about events does not convince philosophers who hold that sometimes facts, properties, or aspects of events are irreducible relata of causal relations (*see* Sanford, 1985). Some philosophers who concentrate on questions of agency and freedom entertain views of *agent causation*: in human action a person is an irreducible cause (*see* ACTION THEORY). Although Lucy's putting on her

shoes involves many instances of event causation, the ultimate cause of Lucy's shoes being put on is Lucy herself.

Causation is the transfer of something from cause to effect. This is a controversial view. In one version of this view, causation transfers some quantity subject to a conservation law of physics. Hans REICHENBACH propounded and Wesley SALMON developed another version in terms the mark transmission of a "mark", a modification that satisfies certain requirements. The transmission of a mark between processes is a transmission of structure. There are clear positive instances of this view. One controversy involves the generalization of these instances. Another questions whether the application of a notion such as "mark" requires some prior causal commitment.

There is no element of genuine a priori reasoning in causal inference. This is a majority view. Most philosophers believe that Hume refuted the rationalists (*see* RATIONALISM) before him (such as SPINOZA, DESCARTES and, on this issue, HOBBES) and the idealists after him (such as MCTAGGART and BLANSHARD) who hold that causation is intrinsically intelligible. Given a determinate event, according to Hume, anything might happen next, so far as reason and logic are concerned. "The contrary of every matter of fact is still possible; because it can never imply a contradiction" (Hume, 1748, p. 25). Cause and effect are distinct existences, and "the mind never perceives any real connexion among distinct existences" (Hume, 1739, p. 636). Reason by itself cannot predict what will happen next after one billiard ball bumps into another. But from what should one attempt to make such predictions, from descriptions of the events in question? If so, which logical relations do or do not obtain will depend on the nature of the description. Any event has logically independent descriptions, and any two events have descriptions that are not logically independent (*see* Davidson, 1980, essay 1). The view that there is at least sometimes an intelligible connection between cause and effect does not rely on inventing clever descriptions. Rather, it concedes a lot to Hume without conceding everything. Just from observing its sensible qualities, we cannot figure out a thing's causal capacities. And when we do come to believe, from a much broader experience, what they are, our evidence does not entail our conclusion. It is still logically possible that anything will happen next. Our beliefs about the physical properties of belts and pulleys are fallible and based on more than an initial visual impression. Still, given the physical properties of the belt and pulley, the spatial relations between them, and the assumption that the belt moves in a certain direction, one can figure out which way the pulley rotates. Although one can draw on experience of similar set-ups that involve belts and pulleys when closing the final gap of causal inference, it is unnecessary to do so. Reason can bridge the gap unaided by additional experience (*see* Sanford, 1994).

By the very nature of causation, effects are never earlier than their causes. This is a majority view. Mackie (1974, ch. 7) discusses the conceptual possibility of "backward causation" and provides further references. There are also serious philosophical discussions of the conceptual possibility of "time travel" in which in there are closed causal loops (*see* LEWIS, 1986).

By the very nature of causation, causes are always earlier than their effects. This is a controversial view. Other requirements of causal connection are symmetric in form; they do not distinguish effects from causes. Defining causal priority in terms of temporal priority thus has theoretical appeal. But there is also a theoretical drawback: the equally appealing account of temporal priority by reference to causation will be circular if the explanation of causal priority is to be temporal. Moreover, simultaneous causation appears not only to be possible, but actual. Physics assures us that much of this appearance is illusion. Since nothing transmits motion faster than the speed of light, the motion of one's fingers, that grip the handle of a teaspoon, does not, strictly speaking, cause the simultaneous motion of the bowl of the spoon. Other cases of apparent simultaneous causation, however, do not involve bridging a spatial gap, as when

a moving belt turns a pulley with which it is in direct contact.

We cannot directly perceive causal relations. This is a majority view that Hume influences greatly with his example of the impact of billiard balls. We can see motions and changes in motion in the balls. We can see that one ball touches the other immediately before the second begins to move. We cannot see that there is a causal relation between the two motions. Nor can we tell, just by observing the sensible qualities of a thing, what are its causal capacities and dispositions.

Our sense of touch and our perceptions of the positions and movements of our limbs enable our direct perception of causal relations (*see* von WRIGHT, 1971, pp. 66–74). This is a minority view. The causal relations between one's arm movement and the movement of a cue stick one grasps is a more promising candidate for an object of direct perception than the impact of billiard balls that is merely seen. The *conceptual fallacy* (here so named) may be tempt one here. This is a mistaken inference of the form that since we cannot conceive of A without having the concept of B, therefore the existence of A requires the existence of B. It views ontological dependence as following from conceptual dependence. Granted the minority view that our conception of causation depends on our conceptions of ourselves as agents who make things happen in the physical world, and as patients affected by occurrences in the physical world, it does not follow that the existence of causation requires the occurrence of such interactions.

Manipulations are causes. This is a prevailing view. Many languages have many verbs for specific manipulations such as *cook*, *shake*, *turn*, and *hold* that we understand as causal relations. The view is not strictly a truism since it is inconsistent with seriously held positions such as the following. (a) There really is no physical world; its appearance is an illusion; and from this it follows that there really are no genuine manipulations or physical causal relations. (b) Although there really are physical events, those we commonly but wrongly take as cause–effect pairs are really coincident joint effects of a common cause, such as GOD. Current discussions of causation disregard such views and take it for granted that manipulations are causes.

Causation depends on manipulation; a correct general account of causation is in terms of manipulation. This is a minority view. Just because one might reach this view by means of the conceptual fallacy discussed above, that does nothing to prove it false. When distinguished from a view about relations between concepts, however, the theory must deal, by appeal to analogy or imagination, with causal instances in which humans do not and sometimes cannot actually participate, such as those that involve clusters of galaxies.

A correct general account of causation is in terms of intervention. This is a controversial view, which is currently the center of a robust research program (*see* Woodward, 2003). This program is careful to distinguish its technical term "intervention" from the ordinary term "manipulation". Manipulations are performed by agents. While agents also intervene, some natural processes that involve no agents, directly or indirectly, are also called interventions. On the other hand, the notion of an intervention is explicitly causal. Its descriptions use the notion of a causal path. Not all of the descriptions in the literature are equivalent. Here is one description:

> INT is an *intervention* between two variables X and Y on the same causal path if and only if INT completely determines the value of X; every causal path between INT (or any cause of INT) and Y goes through X; and if there is a causal path between Z and Y that neither includes nor is included by the path between X and Y, INT does not affect Z.

Adding fertilizer does not affect the amounts of water and light, which are relevant variables on causal paths that include the growth of tomatoes. According to this definition of intervention, does the addition of fertilizer then intervene on the causal path between nitrogen level and tomato growth? Weeds complicate the answer to

this question. When fertilizer stimulates weed growth, a better tomato crop requires pulling some weeds and bringing the addition of fertilizer under the general description of intervention may also require it.

Since intervention is a thoroughly causal notion, an interventionist account of a specific causal connection is not reductive in the sense of using only non-causal concepts. This need not render such accounts circular. The use of the notion of intervention to support the presence of a specific connection, such as between nitrogen and growth rate, need not assume its presence to begin with. This accords with the function of experiments. *Experiment* is a thoroughly causal notion, yet we use experiments to confirm and to disconfirm causal hypotheses.

Theorems about interventions have a wide scope in understanding the roles of experiments in various sciences. This is a controversial view. From a precise definition of intervention and some strong assumptions about probabilistic relations between variables, theorists prove theorems about directed causal graphs. (There is no attempt here to summarize these results.) While the theorems themselves are neither trivial nor controversial, there is not a consensus about the manner and scope for their useful application to actual causal processes.

Some generalizations that have no exceptions, and some statements of conditional probability, are causal laws. This is a prevailing view. Some universal laws are not causal because they are mathematical or logical laws. Some universal truths are not laws because they are mere "accidental" regularities. If all swimming birds eat fish, this does not imply that there is a law-like connection between birds" swimming and their eating fish. Finding evidence against an accidental regularity, whether quite surprising, or not at all surprising, does not upset our general theories about the world. Providing a general account of the difference between laws and accidental generalization is a major theoretical undertaking. There are many competing theories about the character of PHYSICAL LAWS, for example, the view that laws are relations between properties or universals.

All physical laws are causal laws. This is a majority view. Some philosophers deny that all laws of nature, for example Newton's first law of motion, are causal laws. Consider a body traveling in a straight line, not changing direction or speeding up or slowing down. Where is the causation? Opinions divide on the adequacy of responses such as "Its motion from B to C is caused by its immediately prior motion from A to B."

Events related as cause and effect, when appropriately described, instantiate a physical law. This is a majority view. These appropriate descriptions typically use concepts different from the ones we ordinarily use in describing the causal transaction. Causation in the everyday world supervenes on causal relations that the fundamental laws of nature directly cover. If such SUPERVENIENCE is universal, there are no causal differences without differences of fundamental properties and spatio-temporal arrangements. A singular causal statement need not entail a law, but it does entail that there is a law that covers, probably as described differently, the events mentioned (*see* Davidson, 1980, essay 7).

Causal attribution and the acceptance of corresponding conditional statements are closely related. This is a prevailing view. Hume connects causation with conditionals in this famous passage:

> Similar objects are always conjoined with similar. Of this we have experience. Suitable 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.* Or in other words, *if the first object had not been, the second never had existed.* (Hume, 1748, p. 76)

What Hume puts "in other words" is scarcely a restatement of what goes before. It nevertheless expresses an important and influential claim, that a cause is necessary for its effect.

> Kate turned the key, and the engine started. But if the engine would have started at that very moment anyway, without Kate's key turn, then Kate's turning the key did not start the engine.

If-then statements about what would have happened if something else had occurred are called COUNTERFACTUALS, *contrary-to-fact* or *subjunctive* conditionals. A conditional of the form "If *a* had not happened, then *b* would not have occurred" says that *a* is necessary for *b*: it is impossible for *b* to occur without *a*. If it is impossible for *a* to occur without *b*, then *a* is *sufficient* for *b*. For example, the downward movement of a lever of the first kind is sufficient for the upward movement of its other end. The necessity of *a* for *b* is often separate from the sufficiency for *a* for *b*; the thesis that a cause is both necessary and sufficient for its effect is quite strong. Events or conditions we single out as causes often are neither necessary nor sufficient for their effects. Adding Bob's Super-Grow fertilizer speeded up the growth of the tomato plants, but it was not really necessary. Other brands would have had the same effect. Just by itself, moreover, it also was not sufficient; for other factors, independent of adding the fertilizer, such as light, water and the absence of large amounts of concentrated sulfuric acid, were also necessary for the quick growth of the plants. We can still use the notions of *necessity* and *sufficiency* to spell out the causal relevance of adding Bob's Super-Grow to the plant's rapid growth. It is presumably an *inus condition* of the growth; that is, it is an *insufficient* but *non-redundant* part of an *unnecessary* but *sufficient* condition of rapid growth (Mackie, 1974, p. 62). Inus conditions involve somewhat complicated counterfactual conditionals. The pair of simpler conditionals that express necessity and sufficiency, "If *a* had not happened, neither would *b*" and "If *a* had happened, then so would *b*" together express *counterfactual dependence*.

Causation can be defined in terms of counterfactual dependence (Lewis, 1986, essays 17 and 21). This is a controversial view. Counterexamples provide one source of controversy. Counterexamples to a claim of the form A=B are in general examples of A that are not B or examples of B that are not A. Lucy threw a stone that broke a bottle. If Lucy had not thrown the stone, however, a stone would have broken the bottle anyway. Dorothy was standing by, ready to throw a stone toward the bottle if Lucy did not. Standby causes, over-determination, prevention, and other examples serve as counterexamples to simple formulations of counterfactual conditional accounts. This leads to formulations that are less simple, which in turn stimulates the invention of examples of increasing complexity, and so on, back and forth. (*See* essays in Collins et al., 2004.) Opinions are divided about where this process is leading.

Replacing the notion of counterfactual dependence with the notion of *influence* results in a counterfactual account that runs more smoothly. This is a minority view. One event influences another when each belongs to a range of similar events and there is a range of true counterfactuals of the form if event c (in the first range) had occurred, then event e (in the other range) would have occurred. A mass hanging on a spring influences its length, which varies systematically with the mass. (Within a certain range of values, the relation between mass and length is *invariant*. Invariance and intervention both figure in causal graph theory.) Adding acid to a base exemplifies causal influence. As more acid is added, more base is neutralized. There is, however, a causal relation in this process that seems not to fit the definition of influence. As more acid is added, it is not until all the base is neutralized that the next drop of acid causes a sudden, large increase in acidity (decrease in pH). It remains to be seen how the influence view accommodates this and similar "tipping point" examples in which a small event produces large effect by upsetting an equilibrium.

Questions of causation, inductive support, laws of nature, and counterfactual conditionals are bound closely together. This is a prevailing view. The following distinctions are closely associated, and any one can explain the others: acceptable vs. unacceptable counterfactual conditionals; laws of nature vs. accidental generalizations; a particular observation's inductively confirming vs. not confirming a hypothesis. Acceptable counterfactual conditionals, but not unacceptable ones, fall under laws (as CHISHOLM and GOODMAN have argued). On the other hand,

laws, but not accidental generalizations, support acceptable counterfactuals. Laws, unlike accidental generalizations, are hypotheses that their instances confirm. These interconnections, although mutually explanatory, are arranged in a tight circle and thus evoke a sense of theoretical uneasiness. Philosophers who aspire to develop a theory of causation attempt to break out of the circle by explaining one distinction in the family without appeal to additional distinctions in the same family. Different theories attempt to break out in different places and also differ in their assignments of explanatory priority. For example, one theory holds that a relation between particulars is causal when it falls under a law, while another holds that a generalization is a law when particular causal relations fall under it. No views prevail about the best way to achieve equilibrium in these theoretical matters concerning causation.

An adequate theory of causation should be in terms of PROBABILITY. This is a controversial view. When an event causes another, the occurrence of the cause often increases the probability of the occurrence of the other. However this is not always so. Attempts of formulate universal generalizations connecting probability with causation run up against examples such as the following (an earlier example with more details): Lucy aims a stone at a bottle. She throws it, and the stone breaks the bottle. Whenever they engage in the sport of throwing stones to break bottles, Dorothy throws a stone if Lucy doesn't. Although Lucy often misses, Dorothy almost never misses. Lucy didn't miss this time, however. Her throw broke the bottle. The probability that the bottle would break if she did not throw (and dead-eye Dorothy threw instead) is nevertheless higher than if she did throw. Qualifications of a probabilistic account can accommodate particular examples such as this one, but then, following a pattern of dialectic common in technical philosophy generally, and specifically with the associated counterfactual accounts of causation, new ingenuous counterexamples are not far behind.

a is necessary for *b* if, and only if, *b* is sufficient for *a*. This is a prevailing view that follows from the above standard explanations of *necessary for* and *sufficient for*. This view does not entail the stronger view that *a* is a *necessary condition* of *b* if, and only if, *b* is a *sufficient condition* of *a*. Causal examples, among others, show that "condition of" is not a symmetric relation. The presence of light, for example, is a causally necessary condition of the growth of tomatoes, which is not in turn a causally sufficient condition for the presence of light. No one attempts to produce light by growing tomatoes. A theory of the direction of conditionship can help account for the direction of causation (Sanford, 1975).

A totality of conditions necessary for an occurrence is jointly sufficient for it. This is a controversial view, and not a logical truth, in the technical sense of *sufficient* spelt out above. There is an ordinary sense of *sufficient*, however, namely "enough, lacks nothing". When everything necessary for *b* obtains, the aggregate is collectively sufficient for *b*'s occurrence, because jointly the members of the aggregate are *enough* – nothing necessary for *b* is missing (*see* ANSCOMBE, 1981, p. 135). It is not a logical contradiction to maintain that an event did not occur even though nothing necessary for its occurrence was missing. This contention runs against the grain of the following controversial view:

Something necessitates every event. This is a controversial view. Although what we call a "cause" often falls far short of being sufficient for its effect, it is common to assume that every effect has some, usually more complicated, sufficient cause. The main issue is not whether some occurrences are totally without causal antecedents, but whether, in the technical sense of *sufficient*, every event has a sufficient cause. If every event has a sufficient cause, and every cause is an event, then a classic version of DETERMINISM is true. Every event is a link on a branching chain of causal necessitation that runs from the beginning to the end of the universe. The occurrence of any event is causally consistent with exactly one set of events causally connectible with it, whether these events are earlier or later.

Modern physics, for example in its treatment of atomic decay, discourages belief in determinism. Definitions that resemble Mackie's definition of an inus condition provide for the possibility of causation without sufficiency: *a* is a *suni condition* of *b*, for example, if there is something *x* such that the disjunction *a* or *x* is a necessary condition of *b*, and *x* is not a necessary condition of *b* (Sanford, 1984, p. 58).

Accounts of specific causal connections often refer to causal mechanism. This is a prevailing view. One of the early truly effective drugs was aspirin. As everyone knows, it relieves pain. What scientists did not know, but for years hoped to find, was the mechanism of aspirin's effect. This goal is different from discovering a more general or more fundamental law. Many scientists try to understand mechanisms rather that find general laws that cover certain phenomena, and this is true not just in medicine, biology, and chemistry, but in many other special sciences.

A general account of causation should refer to causal mechanisms rather than to causal laws. This is a minority view. Although operations of mechanisms, of whatever size, seem generally to involve three-dimensional motions, a general theory of causation as mechanism would want a more detailed account of what a mechanism is. Also, some causal connections are so direct that there seems to be no room for a mediating mechanism. Lucy threw a rock that hit a tree before it reached the wall. The tree interrupted the flight path of the rock. Where should one look for the mechanism of this causal interaction?

In Plato's dialogue "The Euthyphro" Socrates and Euthyphro reach a point where they agree that everything all the gods love is pious and that everything pious all the gods love. Socrates goes on to ask whether all the gods love pious things because they are pious, or whether things are pious because all the gods love them. We may call probing questions of this form *Euthyphro Questions* and proceed to ask them about treatments of causation that aspire to provide reductive accounts. Suppose that some theory is sufficiently refined that both conditionals of these corresponding forms are true: when C causes E, a suitably situated relation R obtains; and when a suitably situated relations R obtains, C causes E. (This formulation is due to L. Paul.) The Euthyphro Question is whether (a) C causes E because R obtains or (b) R obtains because C causes E. A philosophical reductive definition, account, or analysis of causation should hope to give an answer of form (a). Some popular accounts appear to favor answers of form (b). Consider a counterfactual statements and a corresponding causal statement:

> If Kate had not turned the key, the engine would not have started.
> Kate's turning the key caused the engine to start.

It is more natural to say that the conditional is true because turning the key caused the engine to start rather than that turning the key caused the engine to start because the conditional is true. Some conditionals are true because of causal connections; causal connections do not obtain because conditionals are true (*see* Sanford, 2003, chs. 11–14). Similarly, causal connections explain the effectiveness of manipulation rather than the other way around. Causal connections also explain the effectiveness of interventions, although interventionist theory does not represent itself as reductive. Theories in terms of the transfer of something, or in terms of underlying mechanism, whatever their difficulties, promise to give appropriate answers to the Euthyphro Question.

In Book II of the *Physics*, ARISTOTLE discusses four kinds of *aitia* or causes. The present article deals only with *efficient* causes. In the "Second Analogy" of the *Critique of Pure Reason* (1781), KANT argues that all changes conform to the law of cause and effect. In "Of Induction", Book III of *A System of Logic* (1843), J. S. MILL presents experimental methods for establishing causal relevance. In his 1912 lecture, "On the Notion of Cause", RUSSELL claims that the law of causation "is a relic of a bygone age"; but Russell's own theoretical constructions in some later writings depend heavily on causal notions.

BIBLIOGRAPHY

Anscombe, G.E.M.: *Metaphysics and the Philosophy of Mind, Collected Philosophical Papers*, vol. 2 (Minneapolis: University of Minnesota Press, 1981).

Beauchamp, T. and Rosenberg, A.: *Hume and the Problem of Causation* (New York and Oxford: Oxford University Press, 1981).

Collins, J., Paul, L., and Hall, N., ed.: *Counterfactuals and Causation* (Cambridge, MA: MIT Press, 2004)

Davidson, D.: *Essays on Actions and Events* (Oxford: Oxford University Press, 980).

Dowe, P.: "Causal Process," in the *Stanford Encyclopedia of Philosophy.*

Faye, J.: "Backward Causation," in the *Stanford Encyclopedia of Philosophy.*

Hausman, D.M.: *Causal Asymmetries* (Cambridge: Cambridge University Press, 1998).

Hitchcock, C. "Probabilistic Causation," in the *Stanford Encyclopedia of Philosophy.*

Hume, D.: *Enquiry Concerning Human Understanding* (London, 1748): ed. L.A. Selby-Bigge (Oxford: Oxford University Press, 1894); 3rd edn. rev. P.H. Nidditch (Oxford: Oxford University Press, 1975).

Hume, D.: *A Treatise of Human Nature*, Book I (London, 1739); ed. L.A. Selby-Bigge (Oxford: Oxford University Press, 1888); 2nd edn. rev. P.H. Nidditch (Oxford: Oxford University Press, 1978).

Lewis, D.K.: *Philosophical Papers*, vol. II (Oxford: Oxford University Press, 1986).

Mackie, J.L.: *The Cement of the Universe*, 2nd edn. (Oxford: Oxford University Press, 1980; originally published 1974).

Menzies, P.: "Counterfactual Theories of Causation," in the *Stanford Encyclopedia of Philosophy.*

Psillos, S.: *Causation and Explanation* (Chesham, Bucks.: Acumen; Montreal: McGill-Queens University Press, 2002).

Salmon, W.C.: *Scientific Explanation and the Causal Structure of the World*, (Princeton, NJ: Princeton University Press, 1984).

Sanford, D.H.: "Causal Relata," in E. LePore and B. McLaughlin, ed., *Actions and Events* (Oxford and New York: Blackwell, 1985), 282–93.

Sanford, D.H.: "Causation and Intelligibility," *Philosophy* 69 (1994), 55–67.

Sanford, D.H.: "The Direction of Causation and the Direction of Conditionship," *Journal of Philosophy* 73 (1975), 193–207.

Sanford, D.H.: "The Direction of Causation and the Direction of Time," *Midwest Studies in Philosophy* 9 (1984), 53–75.

Sanford, D.H.: *If P, then Q: Conditionals and the Foundations of Reasoning*, 2nd edn. (London: Routledge, 2003; originally published 1989).

Schaffer, J.: "The Metaphysics of Causation," in the *Stanford Encyclopedia of Philosophy.*

The *Stanford Encyclopedia of Philosophy* (http://plato.stanford.edu) is an online resource of substantial entries that typically have helpful bibliographies. Entries undergo periodic revision.

Strawson, G.: *The Secret Connexion: Causation, Realism and David Hume* (Oxford: Oxford University Press, 1989).

Woodward, J.F.: "Causation and Manipulability," in the *Stanford Encyclopedia of Philosophy.*

Woodward, J.F.: *Making Things Happen: A Theory of Causal Explanation* (New York: Oxford University Press, 2003).

Wright, G.H. von: *Explanation and Understanding* (Ithaca, NY: Cornell University Press, 1971).

DAVID H. SANFORD

Fictional Entities

The first question to be addressed about fictional entities is: are there any? The usual grounds given for accepting or rejecting the view that there are fictional entities come from linguistic considerations. We make many different sorts of claims about fictional characters in our literary discussions. How can we account for their apparent truth? Does doing so require that we allow that there are fictional characters we can refer to, or can we offer equally good analyses while denying that there are any fictional entities?

While some have argued that we can offer a better analysis of fictional discourse if we accept that there are fictional characters, others have held that even if that's true, we have metaphysical reasons to deny

the existence of fictional entities. Some have supposed that accepting such entities would involve us in contradictions and so must be avoided at all costs, while others have held that, even if contradiction can be averted, we should refrain from positing fictional entities if at all possible since they would be utterly mysterious, involve us in positing unexplained differences in "kinds of being", or violate reasonable calls to parsimony.

1. Linguistic Considerations

At least four sorts of fictional discourse may be distinguished:

(1) Fictionalizing discourse (discourse *within* works of fiction), e.g., "[Holmes was] the most perfect reasoning and observing machine that the world has seen" in "A Scandal in Bohemia".
(2) Non-existence claims, e.g., "Sherlock Holmes does not exist".
(3) Internal discourse by readers about the content of works of fiction. This may be either intra-fictional (reporting the content of a single work of fiction, e.g., "Holmes solved his first mystery in his college years",) or cross-fictional (comparing the contents of two works of fiction, e.g., "Anna Karenina is smarter than Emma Bovary").
(4) External discourse by readers and critics about the characters *as* fictional characters, e.g., "Holmes is a fictional character", "Hamlet was created by Shakespeare", "The Holmes character was modeled on an actual medical doctor Doyle knew", "Holmes appears in dozens of stories", "Holmes is very famous".

The puzzles for fictional discourse arise because many of the things we want to say about fictional characters seem in conflict with each other: How, for example, could Holmes solve a mystery if he doesn"t exist? How could Hamlet be born to Gertrude if he was created by Shakespeare? Any theory of fiction is obliged to say something about how we can understand these four kinds of claim in ways that resolve their apparent inconsistencies. And any theory of fictional discourse will have import for whether or not we should accept that there are fictional entities we sometimes refer to, and if so, what sorts of thing they are and what is literally true of them.

Given these very different types of fictional discourse, many different approaches have been developed, some of which accept and some of which deny that there are fictional entities. Many of the differences among them may be seen as products of differences in which of the four types of discourse each takes as its primary case and central motivator – though of course all are ultimately obliged to say how we should understand each type of discourse.

Perhaps the most popular approach to fictional discourse has been to deny that there are any fictional entities, and to handle the linguistic evidence by adopting a pretense theory. It is plausible that authors in writing works of fiction (and so writing sentences of type (1)) are not making genuine assertions at all, but rather simply *pretending* to assert things about real people and places (Searle, 1979, p. 65). (Though *see* Martinich and Stroll, 2007, ch. 2, for challenges to this.) Inspired by this observation about discourse of type (1), full-blown pretense theories of fictional discourse (such as that developed by Kendall Walton) treat all four forms of fictional discourse as involving pretense and so as making no genuine reference to fictional entities. Discourse of type (3), on these views, involves readers "playing along" with the pretense "authorized" by the work of fiction, and so pretending that what is stated in works of fiction is true. Claims like "Holmes solved his first mystery in his college years" are "authorized" moves in the game of pretense licenced by the work, which is why we find them more acceptable than parallel claims like "Holmes drove a white Plymouth".

While that extension of the pretense view seems plausible enough, more difficulties arise for handling external discourse and non-existence claims. Walton takes external claims of type (4) to invoke new "ad hoc" "unofficial" games of pretense other than those authorized by the story, where, e.g., we pretend that "there are two kinds

of people: "real" people and "fictional characters" (1990, p. 423), or pretend that authors are like gods in being capable of creation, etc. Even apparently straightforward non-existence claims (type 2) are treated as involving pretense: first invoking a pretense that there is such a character to refer to (using the name "Sherlock Holmes"), and then in the same breath betraying that as *mere* pretense, with the addition of "doesn't exist" (1990, p. 422). The full-blown pretense approach thus seems to implausibly take as pretenseful precisely the (type 2 and type 4) talk about fiction that is designed to step outside of the pretense and speak from the real-world perspective. It also offers contorted and *ad hoc* readings of what seem to be straightforward literal claims (cf. Thomasson, 2003). So while pretense theories do well at addressing internal and fictionalizing discourse, they are much less plausible adopted as across the board approaches – but if we can't adopt them across the board, they can't be used to avoid positing fictional entities.

Various other approaches to fictional discourse have been proposed which don' t rely on taking pretense to be ubiquitous in fictional discourse, yet still avoid accepting that there are fictional entities. The best developed of these is Mark Sainsbury's (2005) negative free logic approach, which takes as its central motivation the truth of claims of type (2): non-existence claims involving fictional names. On the negative free logic view, fictional names are non-referring terms, and all simple sentences using non-referring terms are false. Thus "Holmes exists" is false (as "Holmes" doesn't refer), and so its negation "Holmes doesn't exist" is true (Sainsbury, 2005, p. 195), leaving us with a far simpler and more plausible account of the truth of non-existence claims than pretense views provide. Internal discourse by readers can still be held to be true even though it involves non-referring names, since these claims are plausibly held to be implicitly prefixed with a fiction operator, where "According to the fiction, Holmes solved his first mystery in his college years" may be true even if the simple claim "Holmes solved his first mystery in his college years" would be false. Cross-fictional statements can be handled similarly by taking them to fall in the context of an "agglomerative" story operator that appeals to the total content of the relevant stories, taken together, e.g., "According to (*Anna Karenina* and *Madame Bovary* [taken agglomeratively]), Anna Karenina was more intelligent than Emma Bovary" (Sainsbury, forthcoming).

But like the pretense view, the negative free logic view has more difficulties accounting for the apparent truth of external claims of type (4), since their truth cannot be accounted for by taking them as implicitly reporting what is true according to the fiction. Various *ad hoc* ways of interpreting these claims have been tried, e.g., "Holmes is a fictional character", may be read as reporting that, according to some fiction, Holmes exists (Sainsbury, forthcoming). But given the variety of external claims that must be rewritten in different ways, these remain the biggest thorn in the side of negative free logic theories.

On the other side of the debate are those who argue that we can only or best handle fictional discourse by allowing that there are fictional entities and that at least sometimes our discourse refers to them. But even among those who accept that there are fictional entities there are widespread disagreements about what we should consider them to be and what is literally true of them.

Some realist views about fiction are inspired by the apparent truth of internal claims of type (3), and so take fictional entities to be beings that (in some sense) have the properties the characters of the story are said to have, so that claims like "Holmes solved his first mystery in his college years" is true because there is a fictional entity, Holmes, who in some sense has this property. These views have taken many forms – with some taking the fictional entities to be possible people, others taking them to be Meinongian non-existent objects, and others still taking them to be pure abstract entities such as kinds.

One natural approach inspired by the desire to accommodate the truth of type (3) internal claims is to take fictional characters

to be merely possible people described by the stories. KRIPKE expressed this idea when he wrote "Holmes does not exist, but in other states of affairs, he would have existed" (1963/1971, p. 65). But Kripke himself later (1972, p. 158) rejected this answer, and his rejection of it has generally been taken on board. His grounds for rejecting it come from considerations about reference: the name "Sherlock Holmes" is not a description (which could be fulfilled by various possible individuals); instead, if it refers at all, it picks out the individual to whom the speaker's use of the name bears a historical connection, and it refers to that very individual across all possible worlds. So if there happened to be someone in the actual world who coincidentally was just as Holmes is said to be in the novels, that would not show that he *was* Holmes. Similarly, if there are individuals in other possible worlds who fulfill the descriptions in the books, that does not show that any of them is Holmes. Moreover, since there will be a great many different possible individuals who fulfill the descriptions, it seems there would be no non-arbitrary way of saying *which* of these is Holmes (Kripke, 1972, pp. 157–8).

Given the problems with possibilist views, the most popular realist treatments of fictional entities have been not possibilist but Meinongian and abstractist views. Meinong himself was not interested in fiction *per se*, but rather sought to develop a general theory of the objects of speech and cognition (1904/1960). If there is knowledge, Meinong thought, there must be something known, if there is a judgment, there must be something judged, and so on. So, for example, if we know that the round square is round, there must be something (the round square) of which we know that it is round. Some of these objects of knowledge, however (like the round square) do not exist. Meinongian views thus take seriously the truth of internal (type (3)) sentences like "Holmes solved his first mystery in his college years", and take fictional entities to be the NON-EXISTENT OBJECTS truly described in such sentences – so on these views a fictional entity is the object that (in some sense) has all of the properties ascribed to the character in the relevant work (or works) of fiction.

The simple version of this approach encounters difficulties of the kind that led to RUSSELL'S (1905/1990) criticisms of MEINONG. For the stories ascribe to Holmes not only properties like being a person and solving mysteries, but also properties like existing, in conflict with the apparent truth that Holmes doesn"t exist. Indeed Meinongian theories take non-existence claims of type (2) to be straightforwardly true since, although there *are* the relevant fictional entities, they do not exist. So the Meinongian is in danger of contradiction by taking Holmes and the like both to exist (since Meinongian objects are supposed to have all of the properties ascribed to them) and not to exist (since they are non-existent objects).

The central achievement of neo-Meinongians such as Terence Parsons (1980) and Edward Zalta (1983) has been to show how these contradictions may be avoided. Parsons avoids them by distinguishing two kinds of properties: nuclear properties (like being a man, being a detective, etc.) and extra-nuclear properties (like existing, being possible, etc.). He then holds that only the *nuclear* properties ascribed to the character in the story are actually possessed by the corresponding objects, so we do not have to conclude that Holmes exists. Nonetheless, we do need some way to mark the fact that there may be objects (arguably, like Macbeth's dagger) that *don't* exist according to the stories, as well as objects that (like Macbeth) *are* said to exist. To mark this, Parsons suggests that there are "watered down" nuclear properties corresponding to each extra-nuclear property, so that Holmes does not exist (extra-nuclear) but does have watered-down (nuclear) existence. Zalta (1983), following Ernst Mally, avoids contradiction by a different route: distinguishing two modes of predication: encoding and exemplifying. Fictional entities *encode* all of those properties they are said to have in the stories, but that does not mean that they *exemplify* them. So Holmes encodes existence but exemplifies non-existence, and contradiction is avoided.

A third view along similar lines takes fictional entities to be existing *abstract* objects of some sort rather than to be Meinongian non-existent objects. Nicholas Wolterstorff develops one such view, according to which fictional characters are "not persons of a certain kind, but person-kinds" which do exist (1980, p. 144). On this view, authors do not refer to anyone when they write fictional stories; instead, they delineate a certain kind of person by describing certain sets of characteristics. The fictional character Holmes is not a person, but a certain *kind* of person, or "person-kind', that has essentially within it those properties the work attributes to the character, e.g., being a man, being clever, being a detective. . . . As abstracta, of course kinds can't literally *have* such properties as being clever or solving mysteries – but they can be *defined by* the properties essential within them. So on this view, type (3) claims such as "Holmes solved his first mystery in his college years" are true just in case the properties expressed by the predicate (solving one's first mystery during one's college years) are essential within the person-kind Holmes (1980, p. 159). Many (but not all – see below) of the properties attributed to characters in external discourse, e.g., being famous, appearing in stories, may be properties these abstract person-kinds genuinely *have* rather than properties essential within the kind.

But neither of these strategies helps Wolterstorff cope with (type 2) non-existence claims, for existence is ascribed to Holmes in the stories, and so is essential to that person-kind, and the abstract entity that is that person-kind also exists. Wolterstorff suggests two alternative ways of understanding non-existence claims: either as saying that the relevant person-kind has never been exemplified, or (acknowledging Kripke's point) that the author was not referring to anyone when he used the name in writing the story (1980, p. 161).

Despite their differences, possibilist, neo-Meinongian, and abstractist views are alike in taking most seriously internal (type 3) claims about fictional characters, and as a result they face similar difficulties accounting for the truth of at least some type (4) external claims. Whether fictional entities are taken to be unactualized possibilia, non-existent objects, or abstract kinds, it seems that in any of these cases the work of authors writing stories is completely irrelevant to whether or not there are these fictional entities: the relevant possibilia, non-existent objects, and abstract kinds were "around" just as much before as after acts of authoring, and so we can't take seriously the idea that authors create fictional characters on any of these views. The best these views can do to account for the apparent truth of claims such as "Hamlet was created by Shakespeare" is to say that it is at least true that Shakespeare *described* or *selected* Hamlet from among all the available possibilia, non-existent objects, or abstract kinds and, by writing about that object, made it *fictional.* (Below I will return to discuss some metaphysical difficulties these views also face.)

All of the views canvassed thus far – whether or not they accept that there are fictional entities – face difficulties accounting for the apparent truth of certain external (type 4) sentences. This has inspired several recent theorists to begin by taking this sort of discourse as the focal case – a view that requires accepting that there are fictional characters and that these are created by authors in the process of writing works of fiction. Since they take fictional characters to be products of the creative activities of authors, call these "artifactual" views of fiction.

The phenomenologist Roman INGARDEN suggested something like an artifactual view of fiction in his (1931) *The Literary Work of Art*, where he treats fictional characters (and the literary works in which they appear) as purely intentional objects – objects owing their existence and essence to consciousness. Saul Kripke (apparently independently) suggests that fictional entities are human creations in his unpublished 1973 John Locke lectures. He argues that fictional characters exist in the ordinary concrete world (not another possible world), but they do not exist "automatically" as pure abstracta do. Instead, although they are

"in some sense" abstract entities, they are contingent and exist only given concrete activities of writing or telling stories. John Searle (1979, pp. 71–2) similarly claims that authors, in writing stories and pretending to refer to people, instead create fictional characters to which others can then refer. More recently, artifactual views of fiction have been defended by Schiffer (1996) and Salmon (1998), and developed at length by Thomasson (1999, 2003). (VAN INWAGEN (1977, 1983, 2003) develops a similar view according to which fictional characters are theoretic entities of literary criticism, but he is noncommittal about whether or not they are created.)

Artifactualist theories take external (type 4) claims about fictional characters – e.g., that Holmes is a fictional character created by Arthur Conan Doyle, who modeled Holmes on a medical doctor – to be literally true. On Thomasson's view, fictional characters are abstract artifacts created by authors' activities in writing or telling stories, and dependent for their ongoing existence on those stories (and copies or memories of them). The status of fictional characters as created, dependent, abstracta, she emphasizes, is like that of many social and cultural entities such as laws of state, symphonies, and works of literature themselves: none of them may be identified with any concrete entity, none has a definite spatial location, but all come into existence at a particular time given certain types of human activity.

Most artifactualists, like Searle, take fictional characters to be created by authors pretending to refer to real people and places, and so take fictionalizing (type 1) discourse to involve mere pretended assertions. Artifactualists generally do not take (type 3) internal discourse to state literal truths about properties these fictional entities have; instead, they (like Sainsbury fictional entities) typically read these as shorthand for claims about what is true *according to the fiction* or (following Walton) about what is accepted in games of pretense authorized by the story.

The greatest difficulty for artifactual views arises in handling (type 2) non-existence claims. Various strategies may be used here: denials that Sherlock Holmes exists may be read as denials that there is any such person (Thomasson, 1999, p. 112), or any object answering the descriptions in the stories (van Inwagen, 2003, p. 146). Alternatively, these non-existence claims may be read as noting that past users of the name mistakenly supposed that the name-use chain led back to a baptism rather than a work of fiction (van Inwagen, 2003, pp. 146–7; cf. Thomasson, 2003). If some such solution to the problem of non-existence claims can be shown to be plausible and non *ad hoc*, artifactual theories may offer the best overall way to handle fictional discourse – a way which does require positing fictional entities.

2. Metaphysical Considerations

None the less, many think that we have metaphysical grounds to resist positing fictional entities even if we can offer a somewhat better account of language by accepting that there are such entities and that we sometimes refer to them. These arguments have run in parallel to the developing theories of what fictional entities are.

As we have seen, Russell originally claimed that Meinongian objects were "apt to infringe the law of contradiction" (1905/1990, 205); an objection that kept fictional entities largely undefended for over seventy years. While neo-Meinongians showed how to avoid contradiction, their views were none the less widely rejected for drawing a distinction between what objects *exist* and what objects *there are* (or over which we may quantify) – a distinction many philosophers claim to find incomprehensible (van Inwagen, 2003, pp. 138–42).

Abstractist and possibilist solutions, of course, are more acceptable to those already inclined to accept abstract objects, or possible worlds and the objects in them. But even if one accepts that there are platonistic abstracta or mere possibilia (*see* the extended essay on REALISM AND ANTIREALISM ABOUT ABSTRACT ENTITIES), other problems arise in supposing that fictional characters are among them. As mentioned above, fictional characters are generally thought

to be created, contingent features of the actual world, but neither of these is true of either platonistically conceived abstracta (which are eternal and necessary) or of mere possibilia (which are not created by authors and are merely possible). Moreover, some stories are (intentionally or unintentionally) inconsistent, and so some of their characters can't be treated as possible objects having all the properties ascribed in the story.

Another metaphysical problem that arises for both possibilist and abstractist views comes from the fact that they (like the Meinongian views before them) take the descriptions in works of fiction to determine *which* object we are talking about: the fictional entity is the possible person or abstract entity that has, or has essential within it, all of the properties ascribed to the character in the story. But this leads to problems with the identity conditions for fictional characters (*see* Thomasson, 1999, ch. 5). For these views entail that no fictional character could have had any properties other than those they are ascribed. If the author made even a minor change in the work, so that the character is ascribed so much as one different property (however trivial), she would have written about a different possible person, or delineated a different person-kind. As a result, these views must hold that sequels, parodies, and even revised editions must always include entirely different characters from the original texts – in violation of our standard assumption that an author may change what she says about a given character, and that sequels may describe the further adventures of one and the same character. (Meinongian theories face similar difficulties with handling identity conditions.)

Artifactualist views avoid metaphysical dificulties like these by taking fictional characters (like works of literature themselves) to be created by activities of authors and individuated primarily by their historical origin. The artifactualist typically treats historical continuity – not properties ascribed – as the primary factor for the identity of a fictional character. This leaves open the idea that an author might have described a character somewhat differently than she did, and allows that a later author may ascribe new properties to a preexisting fictional character, provided she is familiar with that character and intends to refer back to it and ascribe it new properties (Thomasson, 1999, pp. 67–9).

None the less, artifactualist views face other metaphysical objections. Although the artifactualist treats fictional characters as created entities, they are also clearly abstract in some sense: though not eternal and necessary like the Platonist's abstracta, they still lack a spatio-temporal location (and are not material) (Thomasson, 1999; *see* also CONCRETE/ABSTRACT). But the very idea that there may be created abstracta strikes some as hard to swallow. As Inwagen puts it "Can there really be abstract things that are made? Some might find it implausible to suppose that even God could literally create an abstract object" (2003, pp. 153–4). Thomasson (1999) addresses these worries by noting that those who accept the existence of such ordinary social and cultural objects as laws, marriages, symphonies, and works of literature themselves are apparently already committed to the existence of created abstracta, so that no special problems arise in accepting created abstracta to account for fictional characters. Of course this "companions in guilt" argument leaves us with two choices: allow that there are abstract artifacts and accept the existence of fictional characters, literary works, laws, etc., or deny the existence of all of these and find some way of paraphrasing talk about the latter entities as well as about fictional characters. But those who would take the latter route should note that even accounting for fictional discourse itself is much more difficult if we cannot make reference to the stories in which they appear.

A final and persistent metaphysical argument against fictional entities is that, since it would be much more parsimonious to deny the existence of fictional characters, we should do so if at all possible. The parsimony argument can be addressed in several ways. First, it is worth noting that even Occam's razor only tells us that "it is vain to do with many what can be done with

fewer" – but if we can provide a *better* account of fictional discourse by accepting fictional entities, the antirealist about fictional entities is not really doing the same thing as the realist, with fewer entities. Second, as Thomasson (1999) notes, it is not obviously more parsimonious to do without fictional characters if we must posit abstract artifacts in some other arena, e.g. to make sense of our talk about novels, symphonies, laws of state, and the like.

The most potentially powerful, though also the most controversial, response to parsimony-based arguments comes from a certain minimalist or "pleonastic" approach to their ontology proposed by Stephen Schiffer (1996). On Schiffer's view, pretenseful uses of a fictional name in works of literature, e.g. "[Holmes was] the most perfect reasoning and observing machine that the world has seen", automatically license us to introduce the singular term "the fictional character Sherlock Holmes" which may then be used in a hypostatizing way in literary discussions. Given those prior pretenseful uses, that singular term is guaranteed to refer to a fictional character. But if all that it takes for fictional names to be guaranteed to refer to characters is that these names be used pretensefully in works of literature, it is not at all clear that someone who accepts that there are pretenseful uses of these names in works of literature but denies that there are fictional characters is genuinely offering a more parsimonious view. Instead, as Thomasson argues (2003), such a person would be only twisting the ordinary rules of use for terms like "fictional character" by artificially inflating the conditions it takes for there to be such characters – not offering a genuinely more parsimonious ontology.

3. Broader Relevance

The question of whether or not we should accept that there are fictional entities – and if so, what sort of thing they are – has been a recurrent topic throughout the history of analytic philosophy because of its broader relevance for a range of other philosophical issues. First, as we have seen in section 1, it has relevance for our theory of language. If we deny that there are fictional entities (and so deny that we ever refer to them), we must explain how we can have true statements involving non-referring terms. If we accept that there are fictional entities, we must explain how we can refer to nonexistent objects (if we take a Meinongian view), merely possible objects, or abstracta (whether Platonist or artifactual) – a task that is especially difficult for causal theories of reference, since none of these entities are obviously a part of the actual causal order.

Issues regarding fictional entities also have broader relevance for work in metaphysics. If artifactualists like Thomasson are correct, then whether or not one accepts that there are fictional characters is closely connected to the issue of whether one accepts other mind-dependent social and cultural objects such as laws and nations, stories and symphonies. Moreover, our stance regarding fictional entities has central relevance for issues of ontological commitment and quantification: If the Meinongian is right, we can quantify over entities that don't exist, and existence must be distinguished from quantification. If the minimalist is right, then the measure of ontological commitment is not whether or not we quantify over the relevant entities – for if we accept that there are authors who use fictional names pretensefully in writing works of fiction, we are already tacitly committed to fictional characters regardless of whether they explicitly quantify over them.

See also the A–Z entry on FICTIONAL TRUTH, OBJECTS, AND CHARACTERS.

BIBLIOGRAPHY

Ingarden, Roman: *The Literary Work of Art*, trans. George Grabowicz (Evanston, IL: Northwestern University Press, 1931).

Kripke, Saul: *Naming and Necessity* (Oxford: Blackwell, 1972).

Kripke, Saul: "Semantical Considerations on Modal Logic," in *Reference and Modality*, ed. Leonard Linsky (Oxford: Oxford University Press, 1971; originally published 1963).

Martinich, A.P. and Stroll, Avrum: *Much Ado about Nonexistence: Fiction and Reference*

(Lanham, MD: Rowman and Littlefield, 2007).
Meinong, Alexius: "On the Theory of Objects," in *Realism and the Background of Phenomenology* ed. Roderick Chisholm (Atascadero, CA: Ridgeview, 1960; originally published 1904).
Parsons, Terence: *Non-existent Objects* (New Haven, CT: Yale University Press, 1980).
Russell, Bertrand: "On Denoting," in *The Philosophy of Language*, 2nd edn. Ed. A.P. Martinich (New York: Oxford University Press, 1990; originally published 1905).
Sainsbury, Mark: *Reference Without Referents* (Oxford: Clarendon Press, 2005).
Sainsbury, Mark: "Serious Uses of Fictional Names" (forthcoming).
Salmon, Nathan: "Nonexistence," *Noûs* 32:3 (1998), 277–319.
Schiffer, Stephen: "Language-Created Language-Independent Entities," *Philosophical Topics* 24:1 (1996), 149–67.
Searle, John: *Expression and Meaning: Studies in the Theory of Speech Acts* (Cambridge: Cambridge University Press, 1979).
Thomasson, Amie L.: *Fiction and Metaphysics* (Cambridge: Cambridge University Press, 1999).
Thomasson, Amie L.: "Speaking of Fictional Characters," *Dialectica* 57:2 (2003), 207–26.
van Inwagen, Peter: "Creatures of Fiction," *American Philosophical Quarterly* 14:4 (1977), 299–308.
van Inwagen, Peter: "Existence, Ontological Commitment, and Fictional Entities," in *The Oxford Handbook of Metaphysics*, ed. Michael Loux and Dean Zimmerman (Oxford: Oxford University Press, 2003).
van Inwagen, Peter: "Fiction and Metaphysics," *Philosophy and Literature* 7 (1983), 67–77.
Walton, Kendall: *Mimesis as Make-Believe* (Cambridge, MA: Harvard University Press, 1990).
Wolterstorff, Nicholas: *Works and Worlds of Art* (Oxford: Clarendon Press, 1980).
Zalta, Edward: *Abstract Objects* (Dordrecht: Reidel, 1983).

AMIE L. THOMASSON

Free Will

The metaphysical "problem of free will" has arisen in history whenever humans have reached a certain stage of self-consciousness about how profoundly the world may influence their behavior in ways unknown to them and beyond their control (Kane, 1996, pp. 95–6). Various authors describe this stage of self-consciousness as the recognition of a conflict between two perspectives we may have on ourselves and our place in the universe (e.g., Nagel, 1986). From a personal or *practical* standpoint, we believe we have free *will* when we view ourselves as agents capable of influencing the world in various ways through our *choices* or *decisions*. When faced with choices or decisions, open alternatives seem to lie before us – a "garden of forking paths" into the future, to use a popular image. We reason and deliberate among these alternatives and choose. We feel (1) it is "up to us" what we choose, and hence how we act; and this means we could have chosen to act otherwise. As ARISTOTLE said, "when acting is 'up to us', so is not acting". This "up-to-us-ness" also suggests that (2) the ultimate sources of our choices, and hence of our actions, lie in us and not outside us in factors beyond our control.

Because of these features, free will is often associated with other valued notions such as moral responsibility, autonomy, genuine creativity, self-control, personal worth or dignity, and genuine desert for deeds or accomplishments (Kane, 1996, ch. 6). These two features of free will also lie behind various *reactive attitudes* that we naturally take toward the behavior of ourselves and others from a personal standpoint (P. STRAWSON, 1963). Gratitude, resentment, admiration, indignation, and other such reactive attitudes seem to depend upon the assumption that the acts for which we feel grateful, resentful, or admiring had their origins in the persons to whom these attitudes are directed. We feel that it was up to them whether they performed those acts or not.

DETERMINISM AND COMPATIBILISM

But something happens to this familiar picture of ourselves and other persons when we view ourselves from various *impersonal*, *objective* or *theoretical* perspectives (Nagel, 1986, p. 110). From such perspectives it may appear that our choices or decisions are not really "up to us", but are determined or necessitated by factors unknown to us and beyond our control. The advent of doctrines of DETERMINISM in the history of philosophy is an indication that this worry has arisen. Doctrines of determinism have taken many forms. People have wondered whether their actions might be determined by Fate or by GOD, by the laws of physics or the laws of logic, by heredity or environment, by unconscious motives or hidden controllers, psychological or social conditioning, and so on. There is a core idea running through all these historical doctrines of determinism that shows why they are a threat to free will. All doctrines of determinism – whether logical, theological, physical, biological, psychological or social – imply that at any time, given the past and the laws of nature (*see* LAW OF NATURE) and of logic, there is only one possible future. Whatever happens is therefore inevitable or necessary (it cannot but occur), given the past and the laws.

Doctrines of determinism thus seem to threaten both features of free will mentioned earlier. If determinism is true, it seems that it would not be (1) "up to" agents what they chose from an array of alternative possibilities, since only one alternative future would be possible, given the past and laws. It also seems that, if determinism were true, the (2) sources or origins of choices and actions would not be in the agents themselves, but in something outside their control that determined their choices and actions (such as the decrees of fate, the foreordaining acts of God, heredity and upbringing or social conditioning).

Those who believe, for these or other reasons, that free will and determinism are not compatible – and hence that free will could not exist in a completely determined world – are *incompatibilists* about free will. The opposite view is taken by *compatibilists*, who hold that, despite appearances to the contrary, determinism poses no threat to free will, or at least to any free will "worth wanting" (Dennett, 1984).

Compatibilists characteristically argue that all the freedoms we recognize and desire in ordinary life – e.g., freedoms from coercion or compulsion, from physical restraint, from addictions and political oppression – are really compatible with determinism. Even if the world should be deterministic, they argue, there would still be an important difference between persons who are free from constraints on their freedom of action (such as coercion, compulsion, addiction and oppression) and persons who are not free from such constraints; and we would prefer to be free from such constraints rather than not, even in a determined world. Compatibilism was espoused by some ancient philosophers, such as the STOICS, and also by Aristotle, according to some scholars. But it became especially influential in the modern era, defended in one form or another by philosophers such as HOBBES, LOCKE, HUME, and MILL, who saw compatibilism as a way of reconciling ordinary experience of being free with modern science. Compatibilism remains popular among philosophers and scientists today for similar reasons. By contrast, incompatibilists of the modern era, such as JAMES, regard compatibilism as a "quagmire of evasion" or a "wretched subterfuge", as KANT called the compatibilism of Hobbes and Hume.

Compatibilists also characteristically warn against confusing determinism with FATALISM, the view that whatever is going to happen, is going to happen, *no matter what we do*. Compatibilists, such as Mill, argue that what we decide and what we do will make a difference to how things turn out, even if determinism should be true. And since we do not know the future, we have to deliberate and try to decide upon the best course of action, whether determinism is true or not.

THE CONSEQUENCE ARGUMENT AND INCOMPATIBILISM

The "Compatibility Question" ("Is free will compatible or incompatible with determinism?") has thus been central to modern debates about free will. And the popularity of compatibilism in the modern era has placed the burden of proof on incompatibilists to show why free will must be incompatible with determinism. Incompatibilists have tried to meet this challenge in various ways. The most widely discussed of their arguments for the incompatibility of free will and determinism in modern philosophy is called the "Consequence Argument". It is stated informally by one of its defenders (VAN INWAGEN, 1983, p. 16) as follows:

> If determinism is true, then our acts are the consequences of the laws of nature and events in the remote past. But it is not up to us what went on before we were born; and neither is it up to us what the laws of nature are. Therefore the consequences of these things (including our own acts) are not up to us.

To say it is not "up to us" what "went on before we were born", or "what the laws of nature are", is to say that there is nothing we can now do to change the past or alter the laws of nature (it is beyond our control). If such things are beyond our control, and our present actions are necessary consequences of the past and laws of nature (as determinism entails), then altering the fact that our present actions occur would appear to be beyond our control as well. In short, *if determinism is true, no one can do otherwise* than he or she actually does; and if free will requires the power to do otherwise, or alternative possibilities, then no one would have free will.

This argument has generated a large critical literature. Each premise and step has been questioned. (For useful summaries of the issues, *see* van Inwagen, 1983; Fischer, 1994; Ekstrom, 2000; Kapitan, in Kane, 2002). Compatibilists have usually challenged the argument in either of two ways. The first challenge comes from "classical compatibilists" (such as Hobbes, Hume, and Mill) who defend *hypothetical* or *conditional* analyses of "can" and "can do otherwise". According to such analyses, to say "we can do otherwise" means that "we would do otherwise, *if* we chose or wanted to do otherwise". If such hypothetical analyses are correct, the conclusion of the Consequence Argument ("if determinism is true, no one can do otherwise" would fail. For, being able to do otherwise would merely entail that one would have done otherwise, if (contrary to fact) one had chosen or wanted to do otherwise, or if the past had been different in some way; and such a claim would be consistent with saying that one's present action was determined by the actual past and laws. Much debate about the compatibility of free will and determinism has thus concerned such hypothetical analyses of "can" and "could have done otherwise" favored by classical compatibilists. Incompatibilists reject hypothetical analyses; and powerful objections have been made against them by J.L. Austin, R.M. CHISHOLM and K. Lehrer, among others. Yet hypothetical analyses continue to be defended by many compatibilists, e.g., DAVIDSON and LEWIS. (For an overview of the debates, *see* Berofsky, in Kane, 2002.)

ALTERNATIVE POSSIBILITIES AND MORAL RESPONSIBILITY

A more radical compatibilist challenge to the Consequence Argument consists in denying altogether the assumption that "free will requires the power to do otherwise, or alternative possibilities". Call this assumption AP (for "alternative possibilities"). If AP is false – if free will does not in fact require the power to do otherwise – then the Consequence Argument, it is argued, would also fail to show that free will and determinism are not compatible. But on what grounds could one deny that free will requires the power to do otherwise? The answer lies in the connection between free will and *moral responsibility*. Freedom of *will* is not just any kind of freedom of *action* or freedom to do what you want. Freedom of will has a special relationship to responsibility or accountability for one's actions. Indeed,

many philosophers actually define free will as that kind of freedom (whatever it may be) that is necessary to confer true moral responsibility (and hence genuine praiseworthiness and blameworthiness) on agents.

The connection between free will and moral responsibility is important because a number of "new compatibilists", including Frankfurt (1969), Dennett (1983), Fischer (1994), and Wallace (1994), have denied that moral responsibility requires the power to do otherwise, or alternative possibilities. They reject a principle that Frankfurt has called the Principle of Alternative Possibilities (PAP): Persons are morally responsible for what they have done, only if they could have done otherwise. If free will is the kind of freedom required for moral responsibility and PAP is false, then AP would be false as well: *Free will* (in the sense required for moral responsibility) would also not require the power to do otherwise or alternative possibilities.

Two kinds of examples have been offered by new compatibilists to show the falsity of PAP. The most widely discussed of these two kinds of examples are called "Frankfurt-style examples", after Harry Frankfurt, who introduced the first such example in 1969. Frankfurt posited a controller named Black, whom we might suppose is a neurosurgeon with direct control over the brain of an agent Jones. Jones faces a choice between doing A (say, voting for a Democrat) and B (voting for a Republican). Black wants Jones to do A, but he does not want to intervene unless he has to. So if Black sees that Jones is going to choose A on his own, Black will not intervene. Only if Black sees that Jones is going to choose B will he intervene in Jones's brain, making Jones choose A. Frankfurt asks us to consider the case where Jones chooses A on his own and Black does not intervene. In such a case, Frankfurt argues, Jones could well have been morally responsible for his choice, since Jones acted on his own and Black did not intervene. Yet Jones could not have done otherwise, since, if he had given any indication of choosing otherwise, Black would not have let him. So it seems that Jones can be morally responsible for his choice even though he could not have done otherwise: and PAP is false.

As with the Consequence Argument, an enormous literature has developed around these Frankfurt-style examples. (Overviews of the literature can be found in Fischer, 1994 and Widerker and McKenna, 2003.) Of many objections that have been made against the use of such examples to undermine PAP, the most discussed objection is one originally made by Kane (1985) and developed independently by Widerker (1995) and Ginet, among others. The objection insists that if some morally responsible (free will) choices are undetermined up to the moment they occur, as incompatibilists require, then a Frankfurt controller like Black could not know in advance which choice the agent Jones was going to make until the choice was actually made. In that case, if the controller did not intervene, the agent would have alternative possibilities; and if the controller did intervene, he would have to do so *in advance* to make the agent choose as he wished. But in that case, the controller would be responsible for the choice, not the agent. To meet this objection, a host of new, more sophisticated, Frankfurt-style examples have been developed in the past decade by David Hunt, Eleonore Stump, Alfred Mele and David Robb, Derk Pereboom, and others. The jury is still out on the efficacy of these new Frankfurt-style examples. (For a discussion of this literature, *see* Widerker and McKenna, 2003.)

HIERARCHICAL THEORIES AND OTHER NEW COMPATIBILIST VIEWS

New compatibilists, such as Frankfurt, have also put forward novel (compatibilist) accounts of free will, according to which free will does not require the power to do otherwise. In a seminal essay, Frankfurt (1971) argues that persons, unlike other animals, "have the capacity for reflective self-evaluation that is manifested in the formation of second-order desires" (p. 7) – desires to have or not to have various first-order desires. Free will and responsibility require that we assess our first-order desires or motives and form "second-order volitions" about which

of our first-order desires should move us to action. Our "wills" – the first-order desires that move us to action – are *free*, according to Frankfurt, when they are in conformity with our second-order volitions, so that we have the *will* (first-order desires) we *want* (second-order desires) *to have* and in that sense we "identify" with our will.

Such a theory of free will is called "hierarchical" for obvious reasons. Classical compatibilism is deficient, according to hierarchical theorists such as Frankfurt, because it gives us only a theory of freedom of *action* (being able to do what we will) without a theory of freedom of *will* in terms of the conformity of first-order motives to higher-order motives (being able, so to speak, to will what we will). Hierarchical theories remain compatibilist, however, since they define free will in terms of a conformity (or "mesh") between desires at different levels without requiring that desires at any level be undetermined.

Other new compatibilist accounts of free will, such as those of Watson (1975) and Wolf (1990), are also "mesh theories", but they reject Frankfurt's hierarchical view. For Watson, the relevant mesh required for free agency is not between higher and lower-order desires, but between an agent's "valuational system" (beliefs about what is good or ought to be done), which has its source in the agent's reason, and the "motivational system" (desires and other motives), which has its source in appetite. Watson thus revives the ancient Platonic opposition between reason and desire – arguing that freedom consists in a certain conformity of desire to reason. Wolf's "reason view" takes this approach in another direction that also has ancient roots. She argues that freedom consists in being able to do the *right* thing for the *right* reasons, which requires in turn the ability to appreciate "the True and the Good". Wolf's theory thus has a stronger normative component than other compatibilist theories.

Other new compatibilist approaches to freedom with a normative component include those of Michael Slote, Paul Benson, and Philip Pettit and Michael Smith. Still other new compatibilist theories, e.g., those of P. Strawson (1962) and Wallace (1994), emphasize the role of "reactive attitudes", such as gratitude, resentment and indignation, in our understanding of freedom and responsibility. (For critical surveys of many of these "new compatibilist" theories, *see* the essays by Haji and Russell, in Kane, 2002).

Another significant new compatibilist approach to free will is *semi-compatibilism*, whose chief defender is Fischer (1994; *see* also Fischer and Ravizza, 1998). Fischer is convinced by Frankfurt-style examples and other considerations that *moral responsibility* does not require alternative possibilities. But he also argues that *freedom* does require forking paths into the future, and hence alternative possibilities; and he is convinced by the Consequence Argument that determinism rules out alternative possibilities. The result of these competing considerations is "*semi*-compatibilism": moral responsibility is compatible with determinism, but freedom (in the sense that requires alternative possibilities) is not compatible with determinism.

HARD DETERMINISM AND HARD INCOMPATIBILISM

Incompatibilists have also put forth new accounts of free will in modern philosophy and new defenses of its incompatibility with determinism. Incompatibilism, however, may take two opposing forms: Incompatibilists who affirm the existence of free will and hence deny the truth of determinism are called *libertarians* in modern free will debates. By contrast, incompatibilists who affirm determinism, and thus deny the existence of free will, have traditionally been called *hard determinists*. Hard determinism will be considered here first and then libertarianism.

Classical hard determinism (as held by d'HOLBACH, for example) consists of three theses: (i) free will (in the strong sense required for ultimate responsibility and desert) is not compatible with determinism; (ii) there is no free will in this strong sense because (iii) all events are determined by natural causes (i.e., determinism is true). Modern skeptics about free will who are

sympathetic to hard determinism, such as Honderich (1988), Pereboom (2001), and G. Strawson (1986) tend to accept theses (i) and (ii), but remain non-committal about (iii) – whether universal determinism is true.

These modern skeptics about free will are aware that, with the advent of quantum physics in the twentieth century, it is far less clear that the physical world is the deterministic system imagined by classical Newtonian physics. In the eighteenth century, with Newtonian physics in mind, LaPlace famously imagined that a superintelligence, knowing all the forces of nature and the exact positions and momenta of particles at any one time, could predict with certainty every future event in the minutest detail.

Today it is customary to distinguish *predictability* or this sort from *determinism*. For it is known that even in some classical physical systems (such as those that exhibit chaotic behavior), future behavior may not be predictable, even though such systems continue to be deterministic. Modern quantum physics complicates this classical picture even further (at least on standard interpretations of it), by insisting that no superintelligence could know the exact positions and momenta of all particles at any moment because the particles do not have both exact positions and momenta at the same time (the Heisenberg uncertainty principle); and hence their future behavior is not predictable *or* determined. Yet issues of determinism and indeterminism in physics remain unsettled because there is continuing controversy about the interpretation of quantum physics and about its metaphysical implications.

As a consequence, modern skeptics about free will usually remain non-committal about the truth of universal determinism (thesis iii), preferring to leave that debate to the physicists. But these modern skeptics about free will continue to hold the first two theses of classical hard determinism, namely that (i) free will – in the "true responsibility-entailing" sense – is incompatible with determinism and that (ii) there is, and can be, no incompatibilist (or libertarian) free will of this true responsibility-entailing kind. One of these modern skeptics about free will, Pereboom calls this successor view to hard determinism, *hard incompatibilism*, which is a useful designation for those who hold theses (i) and (ii), but remain non-committal about thesis (iii).

Why do hard incompatibilists continue to believe that incompatibilist or libertarian free will does not exist, if they are unsure of the truth of universal determinism? There are several reasons. First, while hard incompatibilists remain non-committal about indeterminism in physics generally, they tend to believe that *human behavior* is regular and determined for the most part. If indeterminism does exist in the microphysical world, in the behavior of elementary particles, its macroscopic effects on human behavior, they argue, would be negligible and of no significance for free will. Second, hard incompatibilists are convinced by developments in sciences other than physics – in biology (greater knowledge of genetic influences), neuroscience, psychology, psychiatry, social and behavior sciences – that more of our behavior than previously believed is determined by causes unknown to us. For example, controversial neuroscientific experiments of Libet (2002) and others have led psychologists, such as Wegner (2003), to argue that our familiar experiences of conscious willing may be an "illusion".

New research in the neurosciences in general has had an increasing impact on free will debates. (For discussions of this impact, *see* Walter, 2001; Dennett, 2003; Wegner, 2003; and the essays in Libet et al., 1999. For discussions of the implications of quantum physics and other developments in the physical and behavioral sciences for free will, *see* the essays of Hodgson and Bishop, in Kane, 2002; Earman, 1986; and the essays in Atmanspacher and Bishop, 2004.)

There is a third reason why hard incompatibilists are skeptical of an indeterminist or libertarian free will. They insist that if quantum indeterminism at the micro-level did sometimes have macroscopic effects on the human brain or behavior, such indeterminism would be of "no help" to believers in libertarian free will, since such indeterminism

would not enhance, but would only diminish, freedom and responsibility. Suppose a choice was the result of a quantum jump or other undetermined event in a person's brain, they argue. Such undetermined effects in the brain or body would be unpredictable and occur by chance, like the sudden occurrence of a thought or the spasmodic jerking of an arm – quite the opposite of what we take free and responsible actions to be. Undetermined events in the brain and body would therefore undermine our freedom rather than enhance it.

Hard incompatibilists have also been concerned with the impact their denial of free will would have for morality and the meaning of life. Some of them, such as Honderich and Pereboom, argue that we can still live meaningful lives without the illusion of libertarian free will, though some important "life-hopes" and attitudes would have to change. For example, we could no longer believe that criminal punishment was ultimately deserved. Yet, we could still incarcerate criminals to deter them and others from committing future crimes or to reform them. But other philosophers, such as Smilansky (2000), who also believe libertarian free will is impossible, argue that the effects on society and moral life would be dire if most people became convinced that we do not have an incompatibilist or libertarian free will. Smilansky provocatively suggests that while we do not have such an incompatibilist free will, we must continue to foster the illusion which most ordinary persons share that we do have such a free will for the sake of morality and social order.

LIBERTARIAN VIEWS OF FREE WILL

Libertarianism is the name usually given to those who hold that (i) free will and determinism are incompatible, (ii′) free will (in this incompatibilist sense) exists and so (iii′) determinism is false. Libertarianism about free will in this sense should not be confused with the political doctrine of libertarianism, the view that governments should be limited to protecting the liberties of individuals so long as the individuals do not interfere with the liberties of others. Libertarianism about free will and political libertarianism share a name – from the Latin *liber*, meaning free – and they share an interest in freedom. But libertarians about free will are not necessarily committed to political libertarianism and may (and many do) hold differing political views.

To defend libertarianism about free will, one has to do more than merely argue for the incompatibility of free will and determinism. One must also show that we can actually have a free will that is incompatible with determinism. Many philosophers, including both hard determinists and compatibilists, have argued that an incompatibilist free will of the kind that libertarians affirm is not even possible or intelligible and that it has no place in the modern scientific picture of the world. Critics of libertarianism note that libertarians have often invoked obscure and mysterious forms of agency or causation to defend their view.

In order to explain how free actions can escape the clutches of physical causes and laws of nature, libertarians have sometimes posited a disembodied mind or SOUL in the manner of DESCARTES, which is outside of the physical realm and not governed by physical laws, yet capable of influencing physical events. Other libertarians, such as Kant, have appealed to a noumenal self, outside space and time, not subject to scientific causes and explanations. Still other libertarians, such as REID, appeal to a special kind of agent- or immanent causation that is irreducible to ordinary forms of CAUSATION (*see* the extended essay) in terms of events common to the sciences. Appeals such as these, and other appeals by libertarians to uncaused causes or unmoved movers, have invited charges of obscurity or mystery against libertarian views of free will by their opponents. Even some of the greatest defenders of libertarianism, such as Kant, have argued that we need to believe in libertarian freedom to make sense of morality and true responsibility. But we cannot completely understand such a freedom in theoretical and scientific terms.

The problem that usually provokes skepticism about libertarian free will has to do with an ancient dilemma: If free will is not

compatible with determinism, it does not seem to be compatible with *indeterminism* either. Events that are undetermined, such as quantum jumps in atoms, happen merely by chance. So if free actions must be undetermined, as libertarians claim, it seems that they too would happen by chance. But how can chance events be free and responsible actions? To defend their view, libertarians must not only show that free will is incompatible with determinism, they must also show how free will can be *compatible* with *indeterminism*.

Libertarian accounts of free will have taken a number of different forms in the attempt to address this problem. It is now customary to distinguish three main types of libertarian theories: (1) non-causalist (or simple indeterminist) views; (2) causal indeterminist or event-causal views; and (3) agent-causal views.

Non-causalist or *simple indeterminist* libertarian views rely on a distinction between two ways of explaining events, explanations in terms of *reasons* and purposes (desires, beliefs and intentions) and explanations in terms of *causes*. Non-causalists, such as Ginet (1990) and McCann (1998), argue that free actions can be explained in terms of the agent's reasons for action (desires, beliefs, etc.), without being caused or determined, because explanations in terms of reasons are not causal explanations. Non-causalist views raise important questions of ACTION THEORY about the nature of action, about the distinction between *actions* and other *events* (*see* EVENT THEORY), about whether *reasons* for action can be *causes* of action, and so on. Critics of non-causalist or simple indeterminist views note that, for non-causalists, free actions are literally *uncaused* events, and the critics raise questions about how events that are uncaused can be under the *control* of agents.

Agent-causalist libertarians follow Reid in postulating a special kind of causation by an agent or substance that does not consist in causation by events or states of affairs, as is common for forms of causation studied by the sciences. Agent-causalists, such as Chisholm and O'Connor (2000), insist against simple indeterminists that, while free actions may be uncaused *by events*, they are not uncaused by *anything*. Free actions are caused by the *agents* themselves in a *sui generis* way that is not reducible to causation by states or events of any kinds involving the agent, physical *or* mental. Other agent-causalists, such as Clarke (2003), allow that reasons for action (such as desires and beliefs) can causally influence choices and actions. But he also postulates a special non-event causation by agents to explain what tips the balance between reasons for one choice or action rather than another. Critics of agent-causal theories, such as Watson, argue that appeals to a special kind of non-event causation by substances are no less mysterious than Kantian appeals to noumenal selves or Cartesian appeals to disembodied minds to explain undetermined free choices. Agent-causalists have attempted to rebut such charges in various ways. (For an overview of the debates *see* the essays of O'Connor et al., in Kane, 2002; Clarke, 2003, ch. 10).

Causal indeterminist or *event-causal* views (the third kind of libertarian theory) are of more recent origin. Such views were first suggested, though not developed in detail, in the 1970s by David WIGGINS and Robert Nozick as an alternative to non-causalist and agent-causal views. The first fully developed causal indeterminist view was that of Kane (1985, 1996). Causal indeterminists attempt to explain undetermined choices without appealing to claims that reasons cannot be causes of actions and without appealing to "extra factors" such as noumenal selves, disembodied minds, or non-event agent causes to explain free actions. Causal indeterminists allow that free actions may be caused by reasons, intentions and other states and processes of the agent, but not deterministically caused. "Undetermined", they point out, need not mean "uncaused": Reasons can cause actions non-deterministically or probabilistically, so that, while libertarian freedom must be indeterminist, it need not be "contra-causal".

Causal indeterminist or event-causal libertarian views come in two varieties. "Deliberative" views (first suggested by Dennett and later developed by Mele, 2006 and Ekstrom, 2000) place the indeterminism

early in the deliberative processes of agents, in the undetermined "coming to mind" of thoughts, memories and other considerations that influence subsequent choice. By contrast, so called "centered" causal indeterminist views (such as that of Kane) insist that indeterminism can in some cases persist right up to the moment of choice itself.

An important criticism of causal indeterminist views of the centered variety is the so-called "luck objection", an objection that has been used against other libertarian views as well, agent-causal and non-causalist. (*See* Mele, 2006 and Haji, 2003 for extended discussions of this objection.) Mele puts the luck objection this way: Suppose John fails to resist the temptation to tell a lie. If his choice to lie is a free act in the libertarian sense then it must have been undetermined up to the moment it was made. This means John could have done otherwise (could have chosen not to lie), given exactly the same past up to that moment (since indeterminism implies "same past, different possible outcomes"). Thus we can imagine a counterpart of John, John*, in an alternative possible world with exactly the same past who did resist temptation and chose not to lie. Mele argues that, since there is nothing about the powers, capacities, states of mind, moral character and so on that is different in the pasts of John and John* right up to the moment they chose that could explain the difference in their choices, then the difference in their choices would have been merely a matter of luck. John* got lucky in attempting to resist temptation, while John did not; and it would not be fair to reward one and punish the other for what was merely a matter of luck. A considerable literature has been generated by this "luck objection". Causal indeterminists and other libertarians have tried to answer it in various ways, but many believe it cannot be answered.

ULTIMATE RESPONSIBILITY

One final topic concerning incompatibilist and libertarian views of free will deserves mention. Most arguments for the incompatibility of free will and determinism, like the Consequence Argument, have appealed to the requirement of alternative possibilities or AP, or branching paths into the future. But a number of modern incompatibilists about free will, have argued that another requirement of free will, a requirement of ultimate responsibility or UR, is as important as AP, perhaps even more important, to debates about the incompatibility of free will and determinism. The basic idea of UR is this: To be ultimately responsible for an action, an agent must be responsible for anything that is a sufficient cause or motive for the action's occurring. If, for example, a choice issues from and can be sufficiently explained by an agent's character and motives (together with background conditions), then to be *ultimately r*esponsible for the choice, the agent must be at least in part responsible by virtue of choices or actions voluntarily performed in the past for having the character and motives he or she now has. Compare Aristotle's claim that if a man is responsible for the wicked acts that flow from his character, he must be responsible for forming the wicked character from which these acts flowed.

The importance of this UR condition was first noted in recent free will debates independently by Kane (1985) and G. Strawson (1986). Both agreed that UR could not be satisfied in a deterministic world, so it provided a further argument for the incompatibility of free will and determinism that did not appeal to AP. But Kane and Strawson disagreed about whether UR was an intelligible or satisfiable condition. Kane, a libertarian, attempted to show that UR could be satisfied. While Strawson, a hard incompatibilist, argued that UR was an unsatisfiable condition since it would either require an impossible infinite regress of past voluntary actions by which we formed our later characters or it would require some initial character-forming acts that were not determined by prior character. Such initial acts would either be determined by something external to the agent or would occur merely by chance. This regress argument, which Strawson called the "Basic Argument", poses a significant challenge to libertarian accounts of free will; and attempts to answer it by

libertarians have also been an important part of current free will debates.

The requirement of ultimate responsibility or UR has played another role in free will debates. Some incompatibilists, now called "source incompatibilists" (including some hard incompatibilists, such as Pereboom, and some libertarians, such as Eleonore Stump and Linda Zagzebski) argue that UR is the primary condition required for an incompatibilist free will and that alternative possibilities (AP) are not required for free will at all. "Source incompatibilists" of this sort are now often distinguished from "leeway incompatibilists", who hold the more traditional view that AP is the primary reason why free will and determinism are incompatible. Disputes between these two views about the comparative importance of UR and AP for free will have thus also become a significant part of modern debates about free will.

BIBLIOGRAPHY

Atmanspacher, Harald and Bishop, Robert, ed.: *Between Chance and Choice: Interdisciplinary Perspectives on Determinism* (Thorverton, Devon: Imprint Academic, 2002).

Clarke, Randolph: *Libertarian Accounts of Free Will* (Oxford: Oxford University Press, 2003).

Dennett, Daniel: *Elbow Room* (Cambridge, MA: MIT Press, 1984).

Dennett, Daniel: *Freedom Evolves* (New York: Vintage Books, 2003).

Earman, John: *A Primer on Determinism* (Dordrecht: Reidel, 1986).

Ekstrom, Laura Waddell: *Free Will: A Philosophical Study* (Boulder, CO: Westview, 2000).

Fischer, John Martin: *The Metaphysics of Free Will: A Study of Control* (Oxford: Blackwell, 1994).

Fischer, John Martin and Ravizza, Mark: *Responsibility and Control: A Theory of Moral Responsibility* (Cambridge: Cambridge University Press, 1998).

Frankfurt, Harry: "Freedom of the Will and the Concept of a Person," *Journal of Philosophy* 68 (1971): 5–20.

Ginet, Carl: *On Action* (Cambridge: Cambridge University Press, 1990).

Haji, Ishtiyaque: *Deontic Morality and Control* (Cambridge: Cambridge University Press, 2002).

Honderich, Ted: *A Theory of Determinism*, 2 vols. (Oxford: Clarendon Press, 1988).

Kane, Robert, ed. *Free Will and Values* (Albany, NY: State University of New York Press, 1985).

Kane, Robert, ed.: *The Oxford Handbook of Free Will* (Oxford and New York: Oxford University Press, 2002).

Kane, Robert, ed.: *The Significance of Free Will* (Oxford: Oxford University Press, 1996).

Libet, B., Freeman, Anthony, and Sutherland, Keith, ed.: *The Volitional Brain: Towards a Neuroscience of Free Will* (Thorverten, Devon: Imprint Academic, 1999).

McCann, Hugh: *The Works of Agency: On Human Action, Will and Freedom*. (Ithaca, NY: Cornell University Press, 1988).

Mele, Alfred: *Free Will and Luck* (Oxford: Oxford University Press, 2006).

Nagel, Thomas: *The View from Nowhere* (New York: Oxford University Press, 1986).

O'Connor, Timothy: *Persons and Causes: The Metaphysics of Free Will* (New York: Oxford University Press, 2000).

Pereboom, Derk: *Living Without Free Will* (Cambridge: Cambridge University Press, 2001).

Smilansky, Saul: *Free Will and Illusion* (Oxford: Clarendon Press, 2000).

Strawson, Galen: *Freedom and Belief* (Oxford: Oxford University Press, 1986).

Strawson, Peter F.: "Freedom and Resentment," *Proceedings of the British Academy* 48 (1962), 1–25.

van Inwagen, Peter: *An Essay on Free Will* (Oxford: Clarendon Press, 1983).

Wallace, R. Jay: *Responsibility and the Moral Sentiments* (Cambridge, MA: Harvard University Press, 1994).

Walter, Henrik: *Neurophilosophy of Free Will* (Cambridge MA: MIT Press, 2001).

Watson, Gary "Free Agency," *Journal of Philosophy* 72 (1975), 205–20.

Widerker, David: "Libertarianism and Frankfurt's Attack on the Principle of Alternative Possibilities," *The Philosophical Review* 104 (1975), 247–61.

Widerker, David and McKenna, Michael, ed.: *Moral Responsibility and Alternative Possibilities* (Aldershot, Hants: Ashgate Publishers, 2003).
Wegner, Daniel: *The Illusion of Conscious Will* (Cambridge MA: MIT Press, 2002).
Wolf, Susan: *Freedom Within Reason* (Oxford: Oxford University Press, 1990).

ROBERT KANE

Individuation

For reasons which will become clear, it is appropriate to begin a general account of individuation with some discussion of *sortal* terms and CONCEPTS. The expression "sortal" is a coinage of John LOCKE's (Locke, 1975, III, III, p. 15). He held a sortal name to signify the complex general idea of a certain sort of things (Locke, 1975, III, VI, p. 1). Prime examples of sortal terms, sometimes also called *substantival general terms*, are "cat", "apple", "mountain", and "star". Sortal terms may be contrasted with *adjectival* terms, such as "red", "round", and "heavy". It is commonly said that the key distinction between sortal and adjectival terms is that while both possess criteria of *application*, only the former possess criteria of IDENTITY (DUMMETT, 1981, pp. 547–8). A criterion of application for a general term tells us *what it applies to*. In other words, it determines the *extension* of the term: the SET of entities all and only the members of which are correctly described by the term, such as the set of *cats* in the case of the sortal term "cat" and the set of *red things* in the case of the adjectival term "red". A criterion of identity for a sortal term tells us what determines whether or not one thing that the term applies to is the same as, or numerically identical with, another thing that the term applies to: whether or not, for instance, the cat that is now sitting on the mat is the same cat as the cat that was formerly sleeping on the sofa. Where "*K*" is a sortal term, the general form of a criterion of identity will be this: If *x* and *y* are *K*s, then *x* is identical with *y* if and only if *x* is R_K-related to *y* (Lowe, 1989b). Here "R_K" denotes a certain *equivalence relation* on *K*s – a relation which must, of course, be distinct from identity itself in order for the criterion in question to be informative and non-circular. (An equivalence relation is one that is reflexive, symmetrical, and transitive.) An adjectival term lacks a criterion of identity because there is no single condition that things to which it applies must satisfy in order to be identical (other than, trivially, identity itself). Thus, there is no such condition that any red thing must satisfy in order to be identical with another red thing: whether or not one red thing is identical with another depends at least in part on what *sort* or *kind* of red things they are – and then the relevant criterion of identity will be that supplied by the relevant sortal term, be it, say, "cat", "apple", or "star".

Sortal *concepts* are what sortal terms express or convey – although, of course, we shouldn't assume that for every sortal concept there exists a sortal term (much less a sortal term in every natural language) which expresses or conveys it. Another name for sortal concepts is *individuative* concepts, for reasons that will become plain when I come, in a moment, to introduce the notion – or, rather, the *notions* – of individuation. But what, quite generally, are *concepts* supposed to be? Of course, this in itself is a highly contentious question. Here I shall simply state one widely held view of the matter, which is that a concept is *a way of thinking of some thing or things* (Lowe, 2006, pp. 85–6). Since thought is a mental process, this means, in effect, that concepts are *mental properties* of a certain kind. For properties or QUALITIES, quite generally, are *ways of being* – ways things are (Lowe, 2006, pp. 14, 90–1). For example, roundness is a way of being shaped and redness is a way of being colored. By the same token, concepts, being ways of thinking of things, are ways of being and hence properties – and, more specifically, *mental* properties, since thought is a mental process. So much for the ontology of concepts. But we speak of thinkers *grasping* or *failing to grasp* concepts. We may take this simply to be a matter of their being able, or not being able, to think of things in certain ways. Someone who grasps the concept of a *cat* is able to think of certain

things – in this case, certain living organisms – in a certain way. What way is that? Well, of course, such a person is able to think of certain living organisms as being *cats*. And what does this involve? Well, among other things, it involves being able to think of these organisms as possessing certain characteristic properties, such as furriness and warmbloodedness, and – most importantly for present purposes – as satisfying a certain criterion of identity. We needn't suppose, however, that a person who grasps the concept of a cat must be able to *articulate* such a criterion in an explicit form, in line with the general form of a criterion of identity stated earlier. Indeed, it is notoriously difficult – even for philosophers – to formulate clear and uncontroversial criteria of identity for many kinds of things, even when we seem to have a good implicit grasp of such criteria that is manifested in our ability to make confident identity-judgments concerning things of those kinds.

So far, we have discussed sortal terms and sortal concepts. In addition, however, there are *sorts* or *kinds*, which sortal terms and concepts purportedly designate. I say "purportedly" for this reason if for no other: even granted that a sortal term or concept *may* designate a really existing sort of things, we can hardly insist that it *must* do so. The point is exactly parallel to one that may be made concerning adjectival terms and concepts or, more generally, predicates and predicative concepts: that they may, but need not, designate anything. For example, it is natural to suppose that "red" denotes a certain color property or quality, *redness*. But, for familiar reasons, it may disputed whether there *really are* any color properties at all. It would, of course, be quite extravagant to suppose that *cats* don't exist, but the history of human thought is replete with examples of sortal terms that failed to designate anything, such as "mermaid", "dragon", "unicorn", and "centaur". What, however, should we say concerning the sortal terms that *do* designate or denote something: *what*, exactly, do they denote? Various answers are possible, one being that they denote, in *plural* fashion, all of the various PARTICULAR things to which they are applicable: so that the sortal term "cat", for instance, denotes *the cats* – all of them – that exist (or, perhaps, all that do, did, or will exist). Another view and more popular view is that a sortal term that has denotation denotes a sort or kind of things conceived as a type of *UNIVERSAL*, which has as its particular instances all of the particular things to which it is applicable. According to this view, the sortal term "cat" denotes a substantial universal or kind, whose particular instances are all the INDIVIDUAL cats that do, did, or will exist (Lowe, 1989a, pp. 157–63).

Now let us turn to another key notion that needs to be clarified for present purposes: that of an *object*. This is a philosophical term of art, which admits of various different interpretations, some narrower than others. In its very broadest use, "object" is interchangeable with the very general all-purpose term "entity". In this sense, anything whatever that does or could exist is an "object", including numbers, properties, propositions, events, surfaces, waves, holes, and places, as well as common-or-garden material objects, such as apples, tables, and rocks. However, I propose to use the term "object" more narrowly to mean an entity that does at least possess *determinate identity conditions* and the kind of unity that makes it something that is, at least in principle, *countable* (Lowe, 1998, pp. 58–61, Lowe, 2006, pp. 75–6). Some of the items listed earlier do not indisputably qualify as objects by this criterion: for example, *waves* do not. In what follows we shall restrict our attention for the most part to *material* objects. However, it is important to recognize that the notion of a material object is still a very broad notion indeed. Crucially, material objects do not collectively constitute a *sort* or *kind* in the sense discussed earlier. In other words, "material object" is not a sortal term and does not express or convey a sortal concept. The reason is simple enough: it is simply not the case that all material objects are governed by the same criterion of identity. Thus, for example, both cats and mountains are material objects, but they do not share the same criterion of identity. All it takes for something to qualify as a material object

is that it (a) be an object, in our narrower sense, and (b) be composed of MATTER. Both cats and mountains qualify by this standard, as do many other material objects governed by yet other criteria of identity, such as tables and stars.

The next important thing to notice is this. Although two different sortal terms, each designating a different sort or kind of things, may convey different criteria of identity for the individual objects to which they apply, this is not necessarily the case and, indeed, is very often not the case. Very often, two such sortal terms convey exactly *the same* criterion of identity. This is the case, for instance, with the sortal terms "cat" and "dog" – and, indeed, with *all* sortal terms denoting kinds of living organism (Lowe, 1998, p. 45). Particulars of all these kinds share the same criterion of identity, which is that of living organisms in general. So it is likewise with all kinds of *material artefact*, for instance, such as tables and computers: they all share the same criterion of identity, which differs from that governing living organisms. But why, it may be asked, must we suppose that all living organisms, say – and certainly all *animals* – share the same criterion of identity? For the following reason. "Animal" – unlike, for instance, "material object" – does at least appear to be a sortal term in good standing, conveying a criterion of identity for the objects to which it applies. After all, we can always intelligibly ask whether an individual animal encountered on one occasion is or is not identical with another individual animal encountered on another occasion – and in order to determine the answer to such a question, we do not necessarily need to know what *sort* or *sorts* of animal these individuals are. Indeed, we may well be uncertain, at least at an early stage of our inquiries, whether we are confronted with just one sort of animal or two, because the individual animals encountered on the two occasions may exhibit very considerable morphological differences, as in the case of a tadpole and a mature frog. However, cats and dogs, say, clearly are both *sorts of animal*. (Indeed, they are clearly *different* sorts of animal.) But in that case the sortal terms "cat" and "animal" must convey *the same* criterion of identity, as must the sortal terms "dog" and "animal", on pain of incoherence. For it is not even *metaphysically* possible that objects of kinds governed by different criteria of identity should be identical (Lowe, 1989a, ch. 4). Hence if the sortal terms "cat" and "animal", say, conveyed different criteria of identity, no individual cat could be identified with any individual animal, which is plainly absurd. But if "cat" and "dog" must, for the foregoing reason, both convey the same criterion of identity as "animal" does, then they must clearly convey the same criterion of identity *as each other* – and the same applies in the case of all other sortal terms denoting animal kinds. This, then, is why I maintain that all animal kinds share the same criterion of identity.

The foregoing discussion, if it is along the right lines, reveals that *general names* – as Locke would have called them – fall into at least three distinct classes. First, there are non-adjectival general terms like "material object" which are certainly *not* sortal terms, because they do not convey any criterion of identity whatever. Second, there are regular sortal terms, such as "cat", "dog", "mountain", "star", and "table", which not only convey a criterion of identity but also purportedly denote certain distinct sorts or kinds. Intermediate in generality between these two classes of general names are non-adjectival general terms like "living organism" and "material artefact", which do convey a criterion of identity but are too general to qualify as regular sortal terms. What these terms designate are not, properly speaking, specific sorts or kinds but, rather, certain *ontological* CATEGORIES – or, more precisely, certain *categories of object* (compare Dummett, 1981, p. 583). What cats and dogs and all other such sorts or kinds have in common is that they are all kinds of *living organism*. The individual members of all these kinds all belong to the same ontological category, the hallmark of this fact precisely being that they are all governed by the same criterion of identity. In effect, then, we can identify those general terms that denote ontological categories – *categorial* terms, as we may aptly call them – as being

the *most general* terms that still convey criteria of identity for the objects to which they apply. And categorial terms fall, in respect of their degree of generality, in between regular sortal terms and *transcategorial* terms, such as "material object". I must emphasize that criteria of identity, on this view, are not empirically discoverable principles, but are, rather, a priori ontological principles which delimit what is and is not metaphysically possible for the objects governed by them (Lowe, 1998, ch. 8).

With this stage-setting in place, we can now at last introduce the term "individuation" itself. This term has two senses (Lowe, 2003). In one sense – which we may call the *cognitive* sense – individuation is a cognitive achievement, consisting in *the singling out of an object in thought* (compare WIGGINS, 2001, pp. 6–7). In this sense, it is *we*, or thinkers quite generally, who individuate objects, whenever we single them out in thought. But in a quite different sense – which we may call the *ontological* sense – individuation has nothing to do with cognition or thinkers, but is simply a certain kind of *metaphysical determination relation between entities*. In this sense, an object is individuated by one or more other entities, its individuator or individuators. An object's individuators, in this ontological sense, are the entities which determine *which object* it is. A simple example drawn from the domain of abstract objects will serve for illustrative purposes. A *set*, then, is individuated, in the ontological sense, by the entities that are its *members*, at least in all cases in which it *has* members (not, thus, in the case of the empty set). If a set has members, its members, and these entities alone, determine *which* set it is. Turning to the case of *material* objects, we can see that material objects of some kinds are individuated by their material parts (at least at some level of decomposition): for example, a heap of stones is individuated by the stones that make it up, because *which* heap it is is determined by which stones make it up. (Note, however, that a heap of stones is *not* individuated by the *subatomic particles* that make it up at any given time, which is why it is important to specify the relevant level of decomposition.) Material objects of some other kinds, however, are apparently not individuated by their material parts (at any level of decomposition). Living organisms seem to be a case in point, for they can undergo a *change* of any of the material parts that they possess at any time during their careers. It is not even clear that – as some philosophers suggest (for instance, KRIPKE 1980) – living organisms are individuated by the material parts that they possess *at their moment of origin*, since it seems that these too could always have been different (Lowe, 1998, pp. 165–6).

It should be clear from these examples that metaphysical principles of individuation are closely related to criteria of identity. But they should not be confused with them. A metaphysical principle of individuation tells us what determines the *identity* of an object, in the sense that it tells us what determines *which* object it is. A criterion of identity, by contrast, tells us what determines whether an object belonging to a given ontological category is or is not *identical with* another such object. In the latter case, we are concerned with identity conceived as a *relation*, whereas in the former case we are concerned with "identity" in the sense of *individual ESSENCE* (to use a traditional term). Identity in this sense, or individual essence, is – as John Locke aptly put it – "the very being of any thing, whereby it is, what it is", this being, according to Locke, the "proper original signification" of the word "essence" (Locke, 1975, III, III, p. 15).

Plausibly, *every* object is individuated in the ontological sense. In every case, *something* – some entity or entities – individuates the object in question, in the sense of determining *which* object it is. To suppose that there are *unindividuated objects* seems incoherent. For an unindividuated object would be an object concerning which there was *no fact of the matter* as to *which* object it was, and it is very hard to see how this could be the case. Now, clearly, objects of different ontological categories are individuated, in the ontological sense, in very different ways. For example, mountains and islands are individuated, at least partly, by their geographical locations. But living organisms and

material artefacts are plainly not. However, the claim that *every* object is individuated might raise in the minds of some critics the worry that an infinite regress is thereby threatened. The thought would be that if every object is individuated and, moreover, is individuated by one or more *objects*, then there is *no end* to individuation and so, perhaps, no object at all really gets individuated. However, there are two ways, at least, to counter this worry. One is to point out that it was implied earlier only that every object is individuated by *some entity or entities*, but not that the entities in question must always themselves be *objects*. Indeed, I said that mountains and islands are partly individuated by their geographical *locations*, but geographical locations are doubtfully *objects* at all and are certainly not material objects. Another point to bear in mind is that nothing said so far implies that objects may never be *self*-individuating. In fact, it is plausible to claim that what we may call *material* SUBSTANCES are indeed self-individuating, including *living organisms*. According to this view, for example, what determines *which* animal a given animal is is nothing other than *that very animal*.

The idea that some objects are self-individuating is certainly far from being absurd (Lowe, 2003). Indeed, in some cases it seems extremely compelling: for instance, in the case of the empty set. For, given that every set is individuated and that sets which have members are individuated purely by those members, we seem to have little option but to say that the empty set individuates *itself*, for it has no members to individuate it in the only way that other sets are individuated. In opposition to this view, it might be suggested that the empty set is in fact individuated by a certain *property* that it alone possesses and possesses necessarily – the property of being the only set that has no members – and that since this property is an entity that is distinct from the empty set itself, that set is not self-individuating. However, this assumes that the predicate "is the only set that has no members", which undoubtedly applies uniquely to the empty set, does indeed denote a certain *property* which that set possesses. But, as has already been noted, we cannot uncritically assume that every predicate denotes a property, if by a *property* we mean some really existing *entity*, be it a universal or a so-called TROPE. There is no obvious reason to suppose that the predicate now in question denotes a property in this sense. That being so, it is hard to see what we can say about the empty set other than that it is self-individuating. *It alone* is the only entity that determines *which* set it is, since nothing else does.

However, it may seem that, because the empty set is an *abstract* object, we can draw few lessons from its case when considering the individuation of *material* objects. But that conclusion would be too hasty. For what makes it plausible to say that the empty set is self-individuating is the fact that it is an object that does not appear to *depend for its identity* on anything other than itself (on this notion of identity-dependence, *see* Lowe, 1998, pp. 147–9). But this also seems to be a characteristic of what we are calling *material substances*, including living organisms. We may take it to be an essential feature of such substances that, even though they are composed of matter, they are capable of changing their material parts and, indeed, *could have been* made up, at any given time, of material parts numerically distinct from those that *actually* make them up at that time. This, if true, is why they do not depend for their identity upon such parts, in the way that something like a heap or pile of stones does. But, given that they do not depend for their identity upon their material parts, it is not clear *what else* they could depend on for their identity, other than simply *themselves*.

Perhaps, in the end, saying that material substances are self-individuating is not so very different from saying, as some metaphysicians do, that they are individuated by their so-called HAECCEITIES or *thisnesses* (Rosenkrantz, 1993). According to this view, what determines *which* animal a given animal, *a*, is is that animal's haecceity – its property of being *that animal* or, in other words, its property of being identical with *a*. But, it may be asked, are there *really* such "properties" as the property of being

identical with *a*? Is the "property" of being identical with *a* really an *entity* that is distinct from *a* itself? Some metaphysicians may find it hard to believe so. But if haecceities are not genuine entities in their own right, it is difficult to see what it can mean to say that animal *a* is individuated by its "property" of being identical with *a*, other than simply to say that *a* individuates *itself* – that *a* itself is the only entity that determines *which* animal *a* is. And this is, I think, a perfectly *coherent* thing to say. Nor should it be supposed that once we say this about *some* objects, we shall be obliged to say it about *all* objects. For we have already seen that there are plenty of objects, such as piles of stones and sets that have members, which are plainly *not* self-individuating. Furthermore, it seems very reasonable to say that at least *some* entities must be self-individuating, on pain of the sort of infinite regress that was mooted earlier. (Thus, if there are haecceities, must not *they* be self-individuating?) So why not say this about material substances, together with, perhaps, other objects such as the empty set? Anyway, let us adopt it as a working assumption in what follows that material substances, including animals and other living organisms, are indeed self-individuating in the ontological sense.

So far, however, I have said very little about individuation in the *cognitive* sense, but this notion too raises important metaphysical issues, concerning the nature of thought. What I did say is that individuation in this sense is a cognitive achievement, consisting in *the singling out of an object in thought* by a thinker, that is, by a person. A *sortalist*, in this connection, is a theorist who maintains that a thinker can successfully single out an object in thought on a given occasion only *as an object of some specific sort*, that is, as falling under or satisfying some specific sortal concept – a concept that the thinker in question must therefore *grasp* and *apply* in individuating that object on that occasion, in the cognitive sense of "individuate". An *anti-sortalist*, correspondingly, is a theorist who denies the foregoing claim. On the face of it, the sortalist thesis as I have just formulated it is clearly too strong. For, it may be urged, a thinker can surely successfully single out an object in thought before having any conception of what *sort* of object it is. For example, in thinking about a particular animal – let's call it Tom – a thinker surely need not be able to single out Tom as being, say, a *cat*, as opposed to a *dog*, or a *pig*. Maybe so. But can a thinker successfully single out in thought a particular animal, such as Tom, without even grasping that Tom *is an animal*, or at least a living organism? Is it possible, for example, for a thinker successfully to single out in thought a particular animal, Tom, while grasping only that Tom *is a material object*? It is hard to see how this can be possible. For, it seems, one cannot successfully single out an object in thought without grasping *which* object it is that one has thus singled out. However, this is the point at which the cognitive and the metaphysical notions of individuation come together in a crucial way. *Which* object a given object is is something that is determined by that object's individuator or individuators and, as we have seen, objects of different types have different types of individuator. Material objects *as such* have no single type of individuator, because material objects as such do not constitute an ontological category but, rather, fall into many diverse ontological categories, such as living organisms, material artefacts, and geological formations.

Turning aside, for a moment, from the case of material objects, consider the following question: can we intelligibly suppose that a thinker could successfully single out in thought an *abstract* object, such as a *set* – for example, the set of prime numbers smaller than 10, {2, 3, 5, 7}, or the set of planets closer to the sun than Jupiter, {Mercury, Venus, Earth, Mars} – without grasping that the object in question is indeed *a set* and thereby grasping its criterion of identity and principle of individuation? For, assuming as we now are that the set in question is non-empty and therefore individuated by its members, how could a thinker know *which* object this set is without grasping *what it is that determines which object it is*, namely, *its members* – something that, it seems, requires the thinker to grasp that *what* this object is is *a set*. But if a thinker

does not know *which* object it is that he is thinking about, how can he really be said to have "singled out that object in thought"? To single out an object in thought is, at the very least, to think something about *that very object*. But how can a thinker's thoughts be said to fasten upon a certain object in particular, as opposed to some other object, if that thinker cannot even be said to know *which* object it is that he is thinking about?

It doesn't appear that matters are fundamentally different in the case of thoughts about material, as opposed to abstract, objects. Accordingly, it is hard to see how a thinker could successfully single out a material object in thought while conceiving of it as nothing more specific than *a material object* – that is, as an object composed of matter. For conceiving of an object in this way would leave entirely open the question of what determined *which* object it is – and yet, without his having a grasp of what a correct answer to that question would be it is hard to see how a thinker could be said to know *which* object he was, supposedly, thinking about. Thus, while we should be happy to allow that a thinker can successfully single out a material object in thought without conceiving of it as belonging to some quite specific *sort* or *kind*, such as the kind *cat*, or the kind *table*, or the kind *mountain*, we should insist that he must grasp, at least implicitly, to what *ontological category* the object in question belongs – such as *living organism*, or *material artefact*, or *geological formation*. This is not at all to imply, of course, that the thinker need be able to have a linguistic command of such categorial terms as these, only that he must have at least an implicit grasp of the relevant *criteria of identity and principles of individuation*. For without such a grasp the thinker cannot really be said to know *what it is* that he is, supposedly, thinking about. And without knowing *that*, he cannot really be said to have singled out an object in thought.

However, it is unlikely that this claim will go entirely unchallenged. One kind of challenge that is likely to be raised against it focuses on the *perceptual* capacities of thinkers. Against sortalists, it is sometimes complained that their position improbably requires us to suppose that thinkers cannot *perceive* objects which do not fall under sortal concepts grasped by them. It is then pointed out that *very frequently* we find ourselves perceiving some object while simply having *no idea at all* what sort of object it is that we are perceiving. This may happen when, for example, an archaic ARTEFACT of unknown purpose is dug up and we ask ourselves, "What on earth *is* this – a drinking vessel, perhaps, or an oil lamp, or something designed to be used in a religious rite?" (compare Campbell, 2002, pp. 70–1). We undoubtedly *see* and *feel* the object, however, and can *talk about it* intelligibly. So is this not a case in which we have managed to *single out the object in thought* but without having a *sortal* conception of it, quite contrary to the sortalist thesis? The first thing that must be said about this type of example is that, of course, we have already conceded that the sortalist thesis is too strong. The most that we should say is that we cannot single out an object in thought without having, at least implicitly, a *categorial* conception of it, and thereby having at least an implicit grasp of the *criterion of identity* that the object satisfies. We could call this the *categorialist* thesis, as opposed to the stronger *sortalist* thesis. The latter is *stronger*, because it implies the former, but the reverse is not the case. Now, in the foregoing archaeological example, no challenge to the categorialist thesis was even threatened, since we were supposing the discoverers of the mysterious object in question to be convinced, at least, that what they had found was a *material artefact* of some kind – and material artefacts constitute an ontological category.

In this context, it is vitally important to distinguish between *thought and perception*. The categorialist thesis is the claim that a thinker cannot successfully *single out an object in thought* without conceiving of that object as falling under a certain ontological category and thereby grasping a corresponding criterion of identity that he conceives it to satisfy. But *perceiving is not thinking* and there is no reason at all why the categorialist should not accept that a person can *perceive* an object without having any conception whatever as to what ontological

category it falls under. Indeed, there are compelling reasons to accept precisely this. For it is evident that many non-human animals *perceive* objects in their immediate environment, even though it would be utterly extravagant to suppose that those animals are capable of *categorizing* those objects ontologically or grasping the relevant criteria of identity for those objects. A dog, for instance, can surely *see its feeding bowl*, without recognizing that *what* it sees is a *material artefact*. But, conceding this, let us then ask: can the dog successfully *single out that object in thought?* Can the dog think about *its feeding bowl – that very object*, as distinct from any other? There seems to be no compelling reason to suppose that it can.

We may conclude now with a final question: what *cognitive* significance, if any, is there in the fact – assuming that it is a fact – that material substances are, in the ontological sense of individuation, *self*-individuating? There seems to be considerable cognitive significance in this fact. For what it apparently implies is that it is sufficient for a thinker to be able to single out such a substance in thought that that thinker should have *perceived* that substance at some time, knowing on that occasion that *what* he was perceiving was an instance of a certain category of material substance, and have retained a memory of this experience. For instance, having *seen* a certain animal, *a*, knowing that *what* I was then seeing was an animal, and remembering this perceptual encounter with *a*, I can subsequently have singular thoughts about *a* – that is to say, I can continue to single out *a* in thought. In other words, I continue to know *which animal a is*, even if I never again have perceptual contact with *a*. To put this another way, I continue to grasp *a*'s *individual essence*. For, we are supposing, what individuates *a*, in the ontological sense, is just *a itself* – it is just *a* itself that determines *which* object *a* is. Hence, my perceptual encounter with *a*, provided that it is informed by a grasp of the category of object to which *a* belongs – and thus a grasp of *a*'s *general* essence, which it shares with all other members of the same category – makes me acquainted with *a*'s individuator. But if one grasps an object's principle of individuation and is also acquainted with the entities which, according to that principle, are its individuators, then one knows *which* object it is. For example, if I grasp the principle of individuation for *sets* and am acquainted with the prime numbers smaller than 10, then I know *which* set, and hence which object, the set of prime numbers smaller than 10 is. What is special about material substances – together, maybe, with some other objects, such as the empty set – is that a thinker does not need to be acquainted with *anything else* in order to be acquainted with such a substance's individuator, so that a grasp of such a substance's general essence together with perceptual acquaintance with that substance provides a thinker with a grasp of that substance's individual essence and thereby an ability to single out that substance in thought, that is, to individuate it in the cognitive sense. But whether this is the *only* way in which a thinker can acquire a grasp of the individual essence of a material substance is another and difficult question.

See also the A–Z entry on INDIVIDUATION.

BIBLIOGRAPHY

Campbell, J.: *Reference and Consciousness* (Oxford: Clarendon Press, 2002).

Dummett, M.A.E.: *Frege: Philosophy of Language*, 2nd edn. (London: Duckworth, 1981).

Kripke, S.A.: *Naming and Necessity* (Oxford: Blackwell, 1980).

Locke, J.: *An Essay Concerning Human Understanding*, ed. P. H. Nidditch (Oxford: Clarendon Press, 1975 [1690]).

Lowe, E.J.: *The Four-Category Ontology: A Metaphysical Foundation for Natural Science* (Oxford: Clarendon Press, 2006).

Lowe, E.J.: "Individuation," in M.J. Loux and D.W. Zimmerman, ed., *The Oxford Handbook of Metaphysics* (Oxford: Oxford University Press, 2003), 75–95.

Lowe, E.J.: *Kinds of Being: A Study of Individuation, Identity and the Logic of Sortal Terms* (Oxford: Blackwell, 1989a).

Lowe, E.J.: *The Possibility of Metaphysics: Substance, Identity, and Time* (Oxford: Clarendon Press, 1998).

Lowe, E.J.: "What Is a Criterion of Identity?," *The Philosophical Quarterly* 39 (1989b), 1–21.

Rosenkrantz, G.S.: *Haecceity: An Ontological Essay* (Dordrecht: Kluwer, 1993).

Wiggins, D.: *Sameness and Substance Renewed* (Cambridge: Cambridge University Press, 2001).

E.J. LOWE

The Mind/Body Problem

Reinhardt Grossmann calls the mind "the great garbage bin of ontology" (1983, p. 256). What seems real but lacks physical respectability we consign to the mind. A long tradition places "secondary qualities" (colors, tastes, sounds, odors) in the mind. These are thought, not to be "out there", but to be "subjective" transitory occurrences in the minds of observers (*see* QUALITY, PRIMARY/SECONDARY). HUME regarded CAUSATION (*see* the extended essay) as a psychological "projection", and it seems natural to distinguish the world as experienced from the world as it is. The idea that minds incorporate non-worldly, non-physical elements, however, evidently places minds outside the physical realm. What science casts asunder, philosophers must piece together. Hence the mind–body problem.

Although he did not invent the mind–body problem, DESCARTES (1596–1650) is responsible for its modern formulation (*see* Matson, 1966). Immediately after proving his existence by noting that the thought expressed by "I exist" must be true if I can so much as consider whether it is true (Meditation 2), Descartes asks, "What am I?" He answers, "a thing that thinks", a thinking SUBSTANCE. Descartes regards planets and trees, not as substances, but as modes, ways extended matter is organized. On the one hand we have extended substance and its modes: material bodies (*see* MATTER). On the other hand, we have thinking substances, minds, and their modes: thoughts, images, feeling (*see* SOUL). Just as minds and bodies are irreconcilable (bodies are extended in space, minds are non-spatial), so modes of thought and modes of extension are incommensurable. Now we are faced with a problem: how could mental goings-on have material effects; how could material occurrences affect the mind? This is Descartes's mind–body problem.

In fact there are two problems here. The first arises from the difficulty of understanding how spatial and non-spatial entities could engage causally. The difficulty is especially pressing for Descartes who regards mental and physical substances as operating on very different laws or principles.

A second difficulty arises from our conception of the physical world as a self-contained closed system. Physical events have, we suppose, purely physical causes. If non-physical minds affect the physical world, it looks as though they would have to initiate or intervene in physical processes. Were that so, the physical world would not be a closed system governed by physical law – a daunting prospect that threatens the garbage-bin status of the mental.

> The self-contained nature of the physical world could be expressed in terms of a conservation principle. Descartes, writing before NEWTON, imagined that what was conserved was motion. Minds could not initiate or inhibit motion in the physical world. Minds could, however, have physical effects without violating physical closure by altering the direction taken by material particles. This solution unraveled with Newton's introduction of force, which moved physics from Cartesian kinematics to a dynamical system. Nowadays we think that what is conserved is mass–energy. In either case Descartes's account of mind–body interaction is no longer viable.

MALEBRANCHE (1638–1715), a Cartesian, sees the problem and rejects interaction. According to Malebranche (and there are suggestions of such a view in Descartes), not only is there no mental–physical causation, there is no purely physical causation. Whatever happens is the result of God's making it the case that mental and physical

substances are as they are at every moment. The world resembles a succession of images on a movie screen. In a movie, events on the screen succeed each other. Their cause, a projector, is not a member of the sequence, however, but something entirely outside it. For Malebranche, GOD does not cause, but "occasions" events in the world. God does this, not by intervening in worldly processes, but making it the case at every instant that a world exists containing those processes (*see* OCCASION, OCCASIONALISM). We should not be shocked by the thought that mental events are causally impotent: physical events are in the same boat!

LEIBNIZ (1646–1716) depicts a world comprising an infinity of independent substances each reflecting the world from a unique point of view. On this conception the physical world amounts to a "virtual world" made up of these points of view. Events unfold in each substance independently but in perfect harmony with events in every other substance. Causal interaction is a harmless illusion.

Both Malebranche and Leibniz skirt the mind–body problem by rejecting mental–physical interaction altogether. If there is no mind–body interaction, there is no mind–body problem. Such maneuvers, however, exact a heavy price. Can we reasonably abandon the idea that physical events are causally connected? Could we ever be satisfied with an account of the world according to which mental occurrences – perceptual experiences, for instance – are not brought about by physical occurrences, and thoughts and decisions never give rise to actions and utterances? Must we settle for the idea that mind–body interaction is illusory?

For Descartes, mental and physical substances are, God aside, mutually exclusive and exhaustive. Each kind of substance has a distinctive attribute: mental substances think, but are not extended; physical substances are extended, but do not think. Mental and physical properties are modes of these attributes, determinate ways be being extended or thinking. Being spherical and being red are ways of being extended. An experience of a spherical red object, in contrast is a mode of thought, a way of being conscious.

Many of Descartes's contemporaries and most of his successors rejected this picture. A mental substance might be a substance with mental properties; a physical substance, one with physical properties. This leaves open another possibility: some substances might have both mental and physical properties, a dualism of properties, not substances. Perhaps mental properties are just distinctive properties of certain complex physical systems.

Property dualism can be developed in various ways. According to Epiphenomenalists – T. H. Huxley (1825–1895), for instance – mental occurrences are by-products of brain processes. They resemble squeaks made by a complex machine that play no role in the machine's operation. When you bark your shin, you feel a pain. This feeling is a result of a chain of events in your nervous system leading from your shin to a region of your brain. In the simplest case, the neurological event that "gives rise to" your painful sensation also produces bodily motions that might otherwise be thought to be caused by the sensation. Conscious states and bodily motions are correlated, not because consciousness is causally efficacious, but because conscious states and bodily motions have common causes.

On the one hand, epiphenomenalism enables us to sidestep worries about mental goings-on intervening in the physical world thereby violating closure. On the other hand, we are left with two significant worries. First, as in the case of Malebranche and Leibniz, we will need to abandon the idea that mentality makes a difference in what we do. You might worry about this, not merely because it seems on the face of it implausible, but because it is hard to see how consciousness could possibly bestow any sort of evolutionary advantage on creatures possessing it. True, consciousness could be an evitable by-product of evolutionarily adaptive physical processes, but it is hard to believe that consciousness itself is evolutionarily irrelevant (Nichols and Grantham, 2000).

A second worry concerns the production of conscious experiences. These are caused by physical processes in the brain, but how is this supposed to work? What exactly is

involved in the production of a non-physical event?

Epiphenomenalists tell us that consciousness "arises from" the brain, but what is this "arising from" relation? Mental events presumably involve mental properties, but where are these properties? They seem not to be among those we discover when we probe the brain. Are they invisible? Are they somehow "outside" space or space–time? The Cartesian problem concerned how extended and non-extended things could interact. The problem arises anew for epiphenomenalism in relation to the production of mental properties or events. The situation appears bleak. We have a robust conviction that, although mental and physical properties are utterly different, interaction between minds and bodies is commonplace. The difficulty is to square this with closure, our conviction that the physical world as a whole is causally closed, mass–energy is conserved. (For a dissenting view, *see* LOWE, 1996.)

One elegant solution is to deny the existence of minds and mental properties altogether. If there are no minds, no mental properties, there is no mind–body problem. HOBBES (1588–1679) argued that we are nothing more than elaborate machines. In a way, Hobbes is just extending Descartes's official view. Descartes held that most human behavior and all behavior of non-human creatures could be explained mechanically. Only in the case of behavior resulting from rational mental processes (most notably linguistic behavior), do we need to posit mental causes. If ratiocination, however, were just a matter of calculation (think of a computing machine to get a feel for what Hobbes has in mind) we would have no need to imagine that our bodies are controlled by minds with distinctive mental properties.

A conception of this kind, materialism, can be developed in two ways (*see* PHYSICALISM/MATERIALISM). First, you might think, as Hobbes does, that mental states and properties are "reducible to", that is identifiable with, physical states and processes. On this view, minds turn out to be brains, mental states and properties turn out to be physical states and properties. Second, you might simply deny that there are minds or mental states or properties (Churchland, 1981; Stich, 1996). To see the difference, consider the discovery of DNA and its consequences for genetics. We now think we can map genes onto complex molecular structures, thereby "reducing" genes to DNA (*see* REDUCTION, REDUCTIONISM). Compare reduction of this kind to the demise of phlogiston. Seventeenth-century chemists explained combustion by supposing that flammable materials contained phlogiston, a fluid driven out when the materials were heated. Advances in chemistry rendered phlogiston superfluous. Phlogiston was not reduced to more fundamental goings-on, but stricken from the scientific inventory. Eliminativists believe a similar fate lies in store for the mind.

According to eliminativists, talk of mental states and properties belongs to an outmoded "folk theory" of human and animal behavior. At one time we explained natural occurrences by supposing objects were animated by spirits. Such explanations were gradually supplanted by explanations adverting exclusively to physical processes. Nevertheless, we persist in regarding human bodies (and the bodies of most animals) as animated by spirits. We comprehend the behavior of intelligent creatures by supposing they are conscious of their surroundings and do what they believe will subserve their interests. Advances in neuroscience, however, promise to undermine "folk psychology" and its posits just as chemical discoveries undermined phlogiston.

You might worry that this way of framing the issues stacks the deck. Consider ordinary beliefs about ordinary objects: tables, trees, volcanoes. Physics and chemistry assure us that these things are at bottom just clouds of particles. We can explain the behavior of these particles without positing the ordinary entities, and there is no prospect of smoothly reducing the ordinary things to respectable physical–chemical kinds. Should we eliminate tables, trees, volcanoes? Mightn't it be better to see talk of tables, trees, and volcanoes as reflecting an inventory of genuine objects that happen to be of no interest to the physicist or chemist?

Physics and chemistry provide us with the deep story about the world, a world that includes the fundamental things and includes as well tables, trees, and volcanoes. These are not add-ons any more that the forest is something in addition to the trees.

Whether or not you are moved by such considerations, even tough-minded philosophers have found eliminativism hard to swallow. We can explain away – "eliminate", consign to the garbage bin – ghosts by supposing that they are illusions, but it is hard to see how this could work with states of consciousness. Illusions seem ineluctably mental. An illusory feeling of pain is still a feeling. Conceiving of mental phenomena as "only in the mind" is scarcely a recipe for their elimination. The problem of reconciling illusions with the physical world is just the mind–body problem all over again.

Materialism dissolves the mind-body problem by subtracting the mental as a distinct category (*see* PHYSICALISM/MATERIALISM). Others, idealists, move in the opposite direction: all that exists are minds and their contents. The physical world is, as George BERKELEY (1685–1753) would put it, a "mere appearance". (For a more recent variant, *see* Foster, 1982.) One advantage of IDEALISM is that it is not hard to see how PHYSICAL OBJECTS could turn out to be illusory. A disadvantage is that idealism addresses the world in a way deeply at odds with tenor, if not the substance, of modern science. The sense is that idealism "works", but only by tossing out the baby with the bath water.

The urge for scientific respectability underlies the advent of psychological behaviorism during the first half of the twentieth century. Behaviorists were intent upon distancing themselves from reliance on introspective techniques to study states of consciousness prominent in the nineteenth century. By their lights this meant providing tough-minded "operational" characterizations of important concepts and shunning anything that might prove objectively unverifiable (Skinner, 1963). The result was psychology minus the mental trappings. Behavior was to be explained by contingencies of "reinforcement" and "operant conditioning". We are conditioned by our involvement with the world to do as we do. The mechanisms are simple but, in combination, yield complex responses.

Meanwhile, philosophers, inspired by WITTGENSTEIN's (1953, §38) insistence that "philosophical problems arise when language goes on holiday", were crafting a philosophical version of behaviorism. Gilbert RYLE campaigned against the "Cartesian myth", the conception of minds as "ghosts in the machine". The mistake, thought Ryle, was to regard mental events as private, inwardly observable goings-on that, while not quite physical, had physical causes (incoming stimuli) and effects (bodily motions). Ryle thought this picture stemmed from a "category mistake" (*see* CATEGORIES): representing "the facts of mental life as if they belonged to one logical type or category . . . when they actually belong to another" (1949, p. 16). A child, watching a parade, is told that a regiment is marching past. Puzzled, the child remarks, "I see soldiers, but where is the regiment?" The child thinks a regiment is something alongside or "over and above" the soldiers, a peculiar sort of object. So it is with us and the mind. Scrutinizing the body, we fail to observe the mind and conclude that minds must be organs like the brain but invisible to outside observers. Rather, Ryle thinks, talk of minds and states of mind is a way of indicating what intelligent agents do or would do under various circumstances. Thoughts and feelings are not inner states. Your thinking of Vienna is just a matter of your being disposed to remark on Vienna or respond with "Vienna" when queried.

Neither Wittgenstein nor Ryle denied that there were inner states, only that states of mind were identifiable with such states. Their aim was to challenge the picture of mental goings on as being causally related to physical goings on. Your forming an intention to stroll does not cause your subsequent strolling. Puzzling over mind–body interaction in such cases manifests a category confusion. Your intention "illuminates" or "makes sense of" your subsequent action. Actions, which presumably have purely physical causes, are understood "in light of"

thoughts and desires. The philosophical mistake is to see these states as ghostly internal causes of behavior.

Despite attempts to move us away from the Cartesian model of minds as inner control centers, philosophical behaviorism came under fire from philosophers who found behaviorist analyses of mental states implausible. Such analyses seek to reduce talk of mental states to talk of behavior or behavioral dispositions (*see* DISPOSITION). If you believe the ice is thin, you will avoid skating on it, or at least be disposed to avoid skating on it – but only assuming that you want not to fall through. Your wanting not to fall through could be analyzed behaviorally, but only by mentioning still further states of mind. What we do or would do depends, it would seem, on interrelations among beliefs and desires, and this resists reductive analysis.

Whatever states of mind are, they do seem to affect behavior causally and to be causally responsive to perceptual inputs from the environment. In the 1950s, U.T. Place (1956) and J.J.C. SMART (1959), colleagues at the University of Adelaide, put forward a mind–brain identity thesis. Mental states, although not analyzable in physical terms, might nevertheless be identified with states of the brain: sensations are brain-processes. This is not something that could be worked out solely by attending introspectively to one's own states of mind, any more than one could work out that lightning is an electrical discharge or that water is H_2O, merely by reflecting on familiar properties of lightning and water. Identities of this kind are discoverable only after careful scientific study. When we investigate the brain, we discover that it has the kind of administrative standing in the processing of incoming stimulation and the production of behavior we associate with the mind. The simplest explanation for this coincidence of roles is that the brain is the mind, mental states are states of the brain.

Plenty of scientists and non-philosophers have thought this for a long time, why not philosophers? Philosophers see the task of reconciling mental and physical properties as fraught with difficulty. The "feel" of a state of mind, its "what-it's-like-ness", its "subjectivity" (Nagel, 1974), seem utterly unlike any physical properties we might hope to discover in the brain. Smart noted that this was so with lightning and electrical discharges, water and H_2O. In both cases properties encountered in experience differed from those we discover via scientific investigation, yet this does not prevent us from identifying lightning and water with electrical discharges and H_2O, respectively. In the case of water and lightning, however, we compare properties of the appearance of water or lightning with properties of the stuff that gives rise to the appearance. In the case of minds and brains, the roles are reversed. What we are trying to explain are the appearances. It would be futile to suggest that we are aware only of the appearances of states of mind.

Philosophical behaviorism succumbed to pressure from the identity theory and transformed itself into functionalism. The stumbling block for psychological behaviorism came with the advent of the computing machine and the Chomskeyean revolution in linguistics. Chomsky (1966) argued that behaviorist categories were hopelessly inadequate to account for human linguistic capacities. At the same time, computing machines were coming to be seen as affording explanatorily tractable models of intelligent behavior. Alan Turing (1950), echoing Hobbes, argued that intelligence could be understood ascomputation. It would be possible in principle to build a mind by programming a machine that would "process symbols" so as to mimic an intelligent human being.

Turing proposed a test for intelligence, the "imitation game". Start with two people, A and B, a man and a woman, communicating via teletype with a third person, the interrogator. The interrogator queries A and B in an effort to determine which is the woman. The woman must answer truthfully, but the man can prevaricate. A wins the game when he convinces the interrogator that he is B. Now, imagine a cleverly programmed digital computer replacing A. If the machine succeeds in fooling the interrogator about as often as a person would,

we should, Turing contends, count it as intelligent.

Despite important advances in technology, events have not born out Turing's optimistic prediction that machines would pass his test by the turn of the century. Still, work in artificial intelligence (AI) has progressed on several, less adventurous fronts. Although attacks on AI (most famously by Hubert Dreyfus, 1972 and John Searle, 1980) have been inconclusive, philosophical enthusiasm for the thesis that the nature of the mind can be captured by a computer program has waned. One question is whether consciousness might supply some needed spark, and this brings us back to the fundamental mind–body problem.

The advent of the digital computer encouraged philosophers to separate what could be called "hardware" questions from questions about "software". Perhaps we should view the mind, not as a physical machine, but more abstractly, as a program running on a physical machine, the brain. What is important is not the mind's physical "implementation", but networks of internal relationships that mediate inputs and outputs. So long as this pattern is preserved, whatever the nature of the underlying "hardware", we have a mind.

This is one way of thinking about FUNCTIONALISM (Fodor, 1968). Functionalists note that we are comfortable ascribing states of mind to very different kinds of physical system. A human being, an octopus, and a Martian could all be said to feel pain, although physical states that might be thought to "realize" pain in each could be very different. This thought led to the thesis that states of mind are "multiply realizable". A property – the pain property, for instance – that has different physical realizers cannot be identified with any of those realizers. This sounds like old-fashioned dualism. But realized properties are realized physically. In this regard they are shaped by, and dependent on, physical goings-on.

Functionalists focus on structure. What matters to a mind is not the medium in which it is embodied (flesh and blood, silicon and metal, ectoplasm), but its organization. Thus construed, functionalism is sometimes traced to ARISTOTLE, who, at times, seemed to be thinking along these lines (*De Anima* Book II, 1–3). One difficulty for any such view is that it seems possible to imagine systems that preserve the same patterns of internal relations as minds, but are not minds. Ned Block (1978) imagines the population of China organized in the way an intelligent system might be organized. Although the Chinese nation is a functional duplicate of a conscious agent, it is hard to think that the nation, as opposed to the individuals who make it up, constitutes a conscious mind.

The functionalist picture is one of "higher-level" mental properties realized by, but distinct from "lower-level" physical realizers. The result is "non-reductive physicalism": minds and their properties are grounded in the physical world, but not reducible to their physical grounds. A similar picture has been inspired by Donald DAVIDSON's "anomalous monism". Davidson (1970) describes the mental as "supervening" on the physical. Davidson borrows the notion of SUPERVENIENCE from R.M. Hare, who had borrowed it from G.E. MOORE. Both Hare and Moore were concerned with issues in ethics. Both, though for different reasons, held that, although moral assertions could not be translated into non-moral, "natural" assertions, moral differences required non-moral differences. If St. Frances is good, an agent indistinguishable from St. Frances in relevant non-moral respects – a "molecular duplicate" of St. Frances – must be good as well. Davidson applied this idea to the relation between mental and physical descriptions: agents alike physically (agents answering to all the same physical descriptions) must be alike mentally (must answer to the very same mental descriptions). Reduction fails – in both ethics and psychology – because agents could be alike morally or mentally, yet differ physically.

Supervenience fits nicely with multiple realization, so nicely that some philosophers began to think of supervenience as providing an account of the realizing relation. Considerable effort was expended on refining the supervenience concept. The result was a proliferation of kinds and grades of supervenience and much discussion as to which

best reflected the relation between mental and physical properties (Kim, 1990). Supervenience, however, is a purely formal, modal notion. If you know that the As supervene on the Bs (moral truths supervene on natural truths, mental truths supervene on physical truths), you know that the Bs in some fashion necessitate the As. But what is responsible for this necessitation? What is it about the Bs that necessitates the As?

There are a number of possibilities: (1) the As are the Bs; (2) the As are made up of the Bs; (3) the Bs include the As as parts; (4) the As are caused by the Bs; (5) the As and the Bs have a common cause. None of these fit what proponents of supervenience or multiple realizability appear to have in mind, however. Sydney SHOEMAKER (1980) has suggested that "causal powers" "bestowed" by mental properties are a subset of powers "bestowed" by a variety of physical realizing properties. When one of these physical properties is on the scene, the mental property is thereby on the scene, option (3) above. Derk Pereboom (2002), invoking the idea that a statue, although "constituted by" a particular lump of bronze, is not identical with the lump, argues that instances of mental properties are wholly constituted by, but not identifiable with their physical realizers, option (2).

These accounts of the realization relation locate mental properties within the physical causal nexus. It is hard to see, however, how any such account could preserve the thought that mental properties are really distinct from their realizers while mingling their causal powers with powers of the realizers. Powers comprising a subset of a thing's physical powers would seem to be physical powers; and powers of a statue are hard to distinguish from powers of the bronze that "constitutes" the statue.

Non-reductive physicalism has proved popular because it promises to preserve the distinctiveness and autonomy of the mental, while anchoring it firmly in the physical world. However, non-reductive physicalism has come under fire from Jaegwon KIM (2005) and others for failing adequately to accommodate mental causation, the centerpiece of the mind–body problem. If mental properties are distinct, higher-level properties, how are they supposed to figure in causal relations involving lower-level physical goings-on? So long as we embrace closure, it appears that physical events – bodily motions, for instance – must have wholly physical causes. The prospect of mental properties making a causal difference in the physical world is evidently inconsistent with mental properties' being irreducible to physical properties and the physical world's being causally closed. We must choose, it seems, between epiphenomenalism – mental properties, although real, are physically impotent – and systematic over-determination – some events have mental causes as well as physically sufficient causes. Kim argues that over-determination is a false option. We thus face a choice between epiphenomenalism, on the one hand, and, on the other hand, the abandonment of the non-reductivist hypothesis. Mental properties are either reducible to physical properties or epiphenomenal. Perhaps, Kim suggests, most mental properties are reducible. Those that are not, qualitative properties of conscious experiences, for instance, the qualia, must be epiphenomenal: real, but causally impotent.

This is close to the line advanced by David Chalmers (1996) in a ringing defense of the irreducible nature of qualia. Chalmers divides mental attributes into those characterizable in "information processing" terms and those that are essentially conscious. The former "logically supervene" on fundamental physical features of organisms: a system with the right sort of functional organization will be intelligent and, in general, psychologically explicable. CONSCIOUSNESS, on the other hand, although determined by the physical facts, is not reducible.

To facilitate the distinction he has in mind, Chalmers imagines zombies, creatures resembling us but altogether lacking in conscious experiences (Kirk, 1974). Such creatures are impossible "in our world", that is, given actual laws of nature. The conceivability of zombies, however, suggests that laws governing the production of conscious qualities are fundamental in the sense that they are additions to laws

governing fundamental physical processes. Think of such laws as analogous to Euclidian axioms. Laws governing consciousness resemble the parallel postulate in being independent of the rest. Their presence or absence has no effect on physical goings-on. Outwardly, a zombie world is indistinguishable from ours.

Both Kim and Chalmers render conscious qualities – qualia – epiphenomenal, perfectly real, but physically irrelevant. The result is what Kim calls "modest physicalism" – physicalism plus a "mental residue" – a conception reminiscent of Descartes's idea that much human behavior is explicable on mechanical principles alone. The difference is that, whereas Descartes embraced interactionism – mental properties are causally potent – Kim and Chalmers regard consciousness as qualitatively remarkable but causally inert.

Other philosophers with physicalist leanings are not so ready to throw in the towel. What exactly are mental qualities, the so-called qualia? Describe a dramatic sensory scene: a sunset viewed from a tropical beach. Your description will invoke a panoply of vivid qualities: colors, odors, sounds. Were we to look inside your head, however, we would observe none of this. Colin McGinn asks "how Technicolor phenomenology could arise from grey soggy matter" (1989, p. 349). As C.D. BROAD reminds us, properties of brains seem utterly different from properties of our conscious experiences.

Let us suppose, for the sake of argument, that whenever it is true to say that I have a sensation of a red patch it is also true to say that a molecular movement of a certain specific kind is going on in a certain part of my brain. There is one sense in which it is plainly nonsensical to attempt to reduce the one to the other. There is something which has the characteristic of being an awareness of a red patch. There is something which has the characteristic of being a molecular movement. It would surely be obvious even to the most "advanced thinker" who ever worked in a physiological laboratory that, whether these "somethings" are the same or different, there are two different characteristics (Broad, 1925, p. 622).

Suppose, however, we distinguish properties of things experienced from properties of experiences. The sunset is red, the breeze balmy, the sand warm, and the waves murmur softly. Colors sounds, odors, and the like are not properties of our experiences of such things, but properties of things we experience, or at any rate properties we represent such things as possessing. The point was made by J.J.C. Smart (1959) in his original discussion of mind–brain identity, and, more recently, others have sought to demystify qualia by arguing that what have been regarded as irreducible qualities of conscious experiences are, in reality, only qualities we represent things as having (Harman, 1990 and Lycan, 1996). Were that so, there would be no insurmountable gulf between mental properties, including properties of conscious experiences, and unexceptional physical properties. Much of the mystery of consciousness might be due to confusion over what experiential properties could be (*see* EXPERIENCE).

Here we have representation playing the garbage-bin role: embarrassing or inconvenient features of the world are consigned to representations of the world. Still, it is difficult to shake the idea that representings are themselves permeated with irreducibly mental qualities. Your being in pain might involve your representing a bodily state as painful, but this representing is, or certainly seems to be, qualitatively loaded.

What we might hope to learn from all this? The mind–body problem takes hold only when we respect the integrity of both the physical and the mental. More often than not this has meant accommodating the mental to the physical, thereby privileging the physical. The ideal solution would involve finding a niche for the mental within the physical realm, but that seems hopeless, no more promising than reduction or elimination. Perhaps we are deluding ourselves. Perhaps we have erred in letting Descartes set the agenda and assuming at the outset that the mental and the physical are mutually exclusive. Suppose, instead, it turned out that the MENTAL/PHYSICAL distinction were not metaphysically deep. In that case, we would have no mystery as

to how mental (in the sense of non-physical) properties could have physical (in the sense of non-mental) causes or effects.

Consider Davidson's "anomalous monism". DAVIDSON is commonly read as holding that mental properties depend on, but are not reducible to physical properties. A mental event is an event with a mental property; a physical event is an event with a physical property. This leaves open the possibility of "token identity" without "type identity": one and the same event could be both mental and physical by virtue of possessing a mental property and a (distinct) physical property. The problem of mental causation arises because we think that events have the effects they have solely in virtue of their physical properties. Mental properties "piggyback" on physical properties, but appear causally inefficacious.

Although this picture is widely attributed to Davidson, it is pretty clearly not what Davidson has in mind. Davidson speaks of descriptions and predicates, not properties. An event is mental, he holds, if it answers to ("satisfies") a mental description; it is physical if it satisfies a physical predicate. One and the same event, including the event's causally efficacious constituent properties, could answer to both a mental and a physical description. For Davidson, the mental–physical distinction is classificatory, not metaphysical. Everything in the world could be given a physical description and so counts as physical. Some portions of the world could also be described using mental terms. TRUTHMAKERS for applications of mental predicates will be fully describable using a physical vocabulary. This is so despite the fact that, owing to very different application conditions, there is no prospect of analyzing mental predicates in physical terms.

A view of this kind treats "mental" and "physical" as classificatory designations, not fundamental metaphysical categories. In this regard it resembles Spinoza's "neutral monism". SPINOZA (1632–1677) held that there is but a single substance possessing multiple "attributes", including the mental and the physical. Finite physical or mental entities are modes of these attributes, ways of being mental or physical. Spinoza's attributes differ from Descartes's, however, in being attributes of a single substance and in being, at a deeper level, unified. In singling out attributes, we are "abstracting" in LOCKE's sense, engaging in "partial consideration" of a substance. Abstraction is a mental act, but what is abstracted is in no way mind-dependent.

These are deep metaphysical waters, but the mind–body problem cries out for a deep solution. Perhaps it is time to abandon the Cartesian presumption that the mental and the physical differ in a fundamental way, along with all the many attempts at reconciliation beholden to the Cartesian presumption. As noted, such attempts have tended to privilege the physical. The mental is seen as reducible to or dependent on the physical in some way. For Davidson and Spinoza, the physical is in no regard privileged. We have one world, variously propertied, describable in various ways, with various degrees of specificity. To imagine that dramatic differences in our modes of classification must reflect fundamental metaphysical discontinuities is to mistake features of our representations of the world for features of the world.

Or so Spinoza and Davidson think. Whether a move to monism represents progress or merely one more philosophical byway leading nowhere remains to be seen. Meanwhile, philosophers will continue to till familiar soil in familiar ways in hopes of bringing forth some new solution to the mind–body problem.

BIBLIOGRAPHY

Block, N.J.: "Troubles with Functionalism," in C.W. Savage, ed., *Perception and Cognition: Issues in the Foundations of Psychology*, Minnesota Studies in the Philosophy of Science 9 (Minneapolis: University of Minnesota Press, 1978), 261–325.

Broad, C.D.: *The Mind and Its Place in Nature* (Paterson, NJ: Littlefield Adams, 1960; originally published 1925).

Chalmers, David: *The Conscious Mind: In Search of a Fundamental Theory* (New York: Oxford University Press, 1996).

Chomsky, N.: *Cartesian Linguistics: A Chapter in the History of Rationalist Thought* (New York: Harper and Row, 1966).
Churchland, P.S.: "Eliminative Materialism and the Propositional Attitudes," *Journal of Philosophy* 78 (1981), 67–90.
Dreyfus, H.L.: *What Computers Can't Do: A Critique of Artificial Reason* (New York: Harper and Row, 1972).
Feyerabend, P.K. and Maxwell, G., ed.: *Mind, Matter and Method: Essays in Philosophy and Science in Honor of Herbert Feigl* (Minneapolis: University of Minnesota Press, 1966).
Fodor, J.A.: *Psychological Explanation: An Introduction to the Philosophy of Psychology* (New York: Random House, 1968).
Foster, J.: *The Case for Idealism* (London: Routledge and Kegan Paul, 1982).
Grossmann, R.: *The Categorial Structure of the World* (Bloomington: Indiana University Press, 1983).
Harman, G.: "The Intrinsic Quality of Experience," *Philosophical Perspectives* 4 (1990), 31–52.
Kim, J.: *Physicalism, or Something Near Enough* (Princeton, NJ: Princeton University Press, 2005).
Kim, J.: "Supervenience as a Philosophical Concept," *Metaphilosophy* 12 (1990), 1–27.
Kirk, R.: "Zombies versus Materialists," *Proceedings of the Aristotelian Society*, suppl. vol. 48 (1974), 135–52.
Lowe, E.J.: *Subjects of Experience* (Cambridge: Cambridge University Press, 1996).
Lycan, W.G.: *Consciousness and Experience* (Cambridge, MA: MIT Press, 1996.)
McGinn, C.: "Can We Solve the Mind–Body Problem?," Mind 98 (1989), 349–66.
Matson, W.I.: "Why Isn't the Mind–Body Problem Ancient?," in Feyerabend and Maxwell (1966), 92–102.
Nagel, T.: "What Is it Like To Be a Bat?," *Philosophical Review* 83 (1974), 435–50.
Nichols, S. and Grantham, T.: "Adaptive Complexity and Phenomenal Consciousness," *Philosophy of Science* 67 (2000), 648–70.
Pereboom, D.: "Robust Nonreductive Materialism," *Journal of Philosophy* 99 (2002), 499–531.
Place, U.T.: "Is Consciousness a Brain Process?," *The British Journal of Psychology* 47 (1956), 44–50.
Ryle, G.: *The Concept of Mind* (London: Hutchinson, 1949).
Searle, J.R.: "Minds, Brains, and Programs," *Behavioral and Brain Sciences* 3 (1980), 417–24.
Shoemaker, S.: "Causality and Properties," in Peter van Inwagen, ed., *Time and Cause* (Dordrecht: Reidel, 1980), 109–35.
Skinner, B.F.: "Behaviorism at Fifty," *Science* 140 (1963), 951–8.
Smart, J.J.C.: "Sensations and Brain Processes," *Philosophical Review* 68 (1959), 141–56.
Stich, S.P.: *Deconstructing the Mind* (New York: Oxford University Press, 1996).
Turing, A.M.: "Computing Machinery and Intelligence," *Mind* 59 (1950), 434–60.
Wittgenstein, L.: *Philosophical Investigations*, trans. G.E.M. Anscombe (Oxford: Basil Blackwell, 1968; originally published 1953).

JOHN HEIL

Modality and Possible Worlds

Propositions are evaluated not only as true or false, but as necessarily or contingently true or false. That seven plus five equals 12 is necessary; that George W. Bush was the President of the United States in 2008 is contingently true, and that Saul Kripke has seven sons is merely possible. What sort of fact makes it true that these propositions have the modal status that they have? The problem is sometimes put in epistemological terms: empiricists, for example, ask how experience could give us reason to believe that a proposition is not just true, but necessary. But the real problem behind this question is not epistemological, and not dependent on any thesis about the sources of our knowledge. Even if an oracle gave us unlimited access to matters of fact about the world, we would still face the question, what could make it the case that some fact was not just true, but *had* to be true?

According to one traditional response to this problem, modal propositions are made true by relations of ideas or linguistic conventions: not by the way the world is, but by the way we conceive or describe it. But (on this view) what is necessary is not that we conceive or describe the world as we do. If it is necessary that all uncles are male, it is not because it is necessary that we should have adopted certain conventions to use the worlds "uncle" and "male" in certain ways. What is said to be a matter of convention – that a certain sentence be used to say something that is true no matter what the facts are – is different from what is said to be necessary, which is the proposition itself that this sentence is conventionally used to express. So how can linguistic conventions, or facts about the way we conceive of things, explain necessity and contingency? In any case, it is hard to see how some statements widely thought to be necessary could be true by convention. How could the way we talk or think make it true and necessary that something (a NUMBER, or GOD, for example) should exist, or that a particular thing (Hillary Clinton, say) should be a member of a particular kind (human being)?

The way we have put the problem is already contentious, since it assumes that the things that are said to be true or false, and necessary or contingent, are *propositions* (*see* PROPOSITION, STATE OF AFFAIRS). An adequate theory of modality must give some account of what propositions are, or of whatever the bearers of truth and necessity are taken to be. One way to begin that is motivated by the empiricist's idea that necessity has its source in relations of ideas or in the meanings of words is with a predicate, not of propositions, but of the sentences of some language (*see* EMPIRICISM). Paradigms of necessary truths, according to this approach, are statements that are logical truths, or truths in virtue of meaning. The significance of this alternative starting point can be illustrated by looking at W.V. QUINE's criticisms of modal logic which began with the assumption that our most basic modal concepts are applied to linguistic expressions, rather than to what they express. Quine distinguished three grades of modal involvement (Quine, 1953). (He was skeptical even of the first, but saw them as increasingly problematic.) The first grade was a necessity predicate of sentences: being logically true, or perhaps being analytic. The second grade was a move from a predicate of sentences to an operator on sentences – from "uncles are male is necessary" to "necessarily, uncles are male". Quine argued that the move involved a use-mention confusion, since operators are to be interpreted in terms of functions whose arguments are the values of expressions, and not the expressions themselves. To stipulate that a sentence of the form "necessarily p" shall be true whenever the sentence that is in the place of "p" satisfies the necessity predicate constrains the interpretation of the operator, but does not determine it. Quine argued that the move from the first to the second grade of modal involvement, while based on a use-mention mistake, was in itself relatively harmless, until one made the further move to the third grade, which was to allow the operator to operate on open as well as closed sentences – that is, to allow quantification into modal contexts. The first move disguised the fact that modal contexts were really quotational, and so that quantification into modal context was, implicitly, quantification into a quotation. One could repair the damage and avoid incoherence, he argued, only by making metaphysical commitments that he and the empiricist developers of modal logic that he was criticizing would agree are unacceptable.

It is true that modern modal logic began (with C.I. Lewis) as a project of analyzing *logical* necessity, and deducibility, so Quine's analysis is appropriate as an *ad hominem* argument against his intended targets (C.I. Lewis and Rudolf CARNAP). But modal concepts in general have much wider application. We may be concerned with what must or might happen, in various senses, and with what would or might have happened under various conditions that did not, in fact, obtain, with the dependence and independence of facts on other facts, and these concerns arise in our attempts to understand and act on the empirical world, and not

just in logic and semantics. To understand modal concepts more generally, it seems appropriate to begin with something like FACTS, states of affairs, or propositions as the things to which modal predicates are applied.

If we begin with a predicate of propositions, rather than sentences, then Quine's three grades of modal involvement look quite different. Suppose we assume, about propositions, only that if we ascribe a well-defined predicate to a determinate entity that is within the range of the predicate, we will have expressed a proposition. Then the move from the first to the second grade of modal involvement looks unproblematic: the proposition expressed by a sentence of the form "necessarily p" will be as well defined as the predicate of propositions with which we began. And the move to the third grade – to an operator on open sentences into which one may quantify – looks unproblematic as well, for the following reason: if an operator on propositions is well-defined, then so is a corresponding operator on propositional functions (functions from individuals to propositions). Suppose the necessity operator, "$\Box$" is interpreted with a function that takes (for example) the proposition that Socrates is human to the proposition that it is necessary that Socrates is human. Suppose that the open sentence "x is human" expresses a function from individuals to propositions. Then "it is necessary that x is human" will express the propositional function whose value, for any individual **a**, is the proposition that the proposition which is the value of "x is human" for argument **a** is necessary.

But even if the move through the grades of involvement is unproblematic, given the assumption that what we start with is a predicate of propositions, a clear account of modality still need an account of propositions (about which Quine was famously skeptical). There are many conceptions of proposition, and lots of controversies about how this notion is best understood, but fortunately we can go some way toward an account of modality while making only minimal assumptions about exactly what propositions are.

Whatever propositions are, all who are willing to talk at all about such things will agree that they have truth conditions, and that their truth conditions are essential to them. Any theory of propositions will say that the class of propositions determines a structure that can be characterized by some familiar interdefinable relations: entailment, incompatibility, consistency, etc. If we start with a notion of consistency or compatibility, as a property of sets of propositions, we can define the other relations that are required in terms of it. We assume that consistency will satisfy the following property: if a set of propositions is consistent, then so is any subset of it. It will be assumed, in a minimal theory of propositions, that every proposition has a *contradictory*, where the notion of a contradictory is definable in terms of the consistency relation as follows: proposition x is a contradictory of proposition y, if and only if, first, the set $\{x,y\}$ is inconsistent, and second, any consistent set of propositions is either consistent with x, or consistent with y. A set of propositions Γ *entails* a proposition x if and only if the set $\Gamma \cup \{y\}$ is inconsistent, where y is a contradictory of x. Two propositions will be *equivalent* if and only if they are mutually entailing. A minimal theory might identify equivalent propositions. Even if finer distinctions between propositions are required for some purposes, we can go some way toward a theory of modality while ignoring such distinctions.

It is clear from the requirements of a minimal theory of propositions that the most basic modal properties are not something added onto a minimal theory of propositions, but are constitutive of it. Intuitively, the consistency of a set is the *possibility* that the members of the set all be true together, and a *necessary* truth is a proposition that is entailed by every set. This is possibility in the widest sense; more restrictive notions of possibility and other modal properties and relations might be defined with additions to the basic structure.

Necessity, according to a familiar slogan going back at least to LEIBNIZ, is truth in all possible worlds, and the notion of a POSSIBLE WORLD has played a prominent role in contemporary treatments of modality, both in formal semantic models, and in the informal

characterization of philosophical problems. (*See* KRIPKE, 1963 for an exposition of the model theory, and Kripke, 1980 for an influential treatments of philosophical problems in metaphysics and the philosophy of language that uses the possible worlds framework.) The notion is a controversial one, and there are substantive disagreements about how it should be understood, and about whether any notion of possible world should play a central role in an account of modality. But at least a minimal concept of possible world can be defined within the basic minimal theory of proposition. Within that theory, we can define *maximal* consistent classes of propositions: classes that are consistent, and that for every proposition contain either that proposition or its contradictory. One might identify a possible (state of the) world with these maximal sets of propositions. Or in an alternative formulation, one might take a set of possible worlds as the primitive basis of one's theory, and define the propositions as sets of them. Whichever primitive notion one begins with, there will be, in a minimal theory, a one-one correspondence between sets of possible worlds and coarse-grained propositions. (*See* Adams, 1974 for an analysis of possible worlds in terms of propositions, and Stalnaker, 2003, ch. 1, for a discussion of the relation between propositions and possible worlds.)

The point of spelling out this minimal theory of propositions, possible worlds and basic modal properties and relations is to set up a framework in which the substantive metaphysical questions about modality can be sharpened and clarified. We will consider questions about the nature of possible worlds and their role in a metaphysical account of modality, but it is useful first to see the minimal framework as an attempt to provides only a paraphrase of problematic modal claims in a language in which ambiguities and equivocations are more easily avoided, and in which the structure of modal claims and questions are more perspicuously displayed. The thesis that necessity is truth in all possible worlds is like Quine's thesis that to be is to be the value of a bound variable. The Quinean thesis is not a substantive claim about ONTOLOGY, but an attempt to get clearer about what such claims come to. I think the thesis that necessity is truth in all possible worlds should be understood in a similar spirit. The paraphrase of modal claims and questions into the language of possible worlds solves some of the more superficial puzzles about referential opacity and merely possible individuals by diagnosing scope ambiguities and by separating questions about names and words from questions about the individuals, kinds and properties that the names and words are used to designate. And it brings to the surface and gives new form to the underlying metaphysical questions about the nature of modality.

In the context of this simple framework, I will consider a number of interrelated metaphysical problems about modality. First, if possible worlds are to be taken as basic entities in our ontology, what kind of thing are they? What is it that makes it true that there are the possible worlds that there are? Different philosophers who take possible worlds to be fundamental to an explanation of modality give radically different answers to these questions. David LEWIS argued that we should take other possible worlds literally as concrete particular universes, spatio-temporally disconnected from our own (Lewis, 1986). Most other philosophers who take possible worlds seriously explain them as possible *states* of the world, ways the world might be. (*See* Kripke, 1980; PLANTINGA, 2003; Stalnaker, 2003 for actualist accounts. *See* Divers, 2002 for a survey of a range of accounts of possible worlds.) These contrasting answers take on different explanatory burdens and give different response to various more specific problems about modality. Second, can we give an account of modality that is *reductive* in some sense, and if so, what is being reduced to what? David Lewis argued that his realist analysis of modality in terms of possible worlds was a reduction of modal to non-modal notions, but others have disputed this. Alternatively, one might try to reduce the notion of a possible world to something more basic. Is a reduction of modal to non-modal notions something we should

seek, or take to be a benefit of a theory if it succeeds, in its own terms, in giving one? Third, might there have been things that do not in fact exist? If so, what can be said about what is merely possible? What ontological commitments are required to make sense of the possibility of things that do not actually exist? The Lewisian modal realist has no problem here (at least no *new* problem), but the actualist needs either to explain what we are really talking about when we seem to be talking about things that might, but don't exist, or else to reject the thesis that there might have been things other than those there are. Fourth, whether or not there might have been things that do not actually exist, it seems obvious that the things there are might have been different in various ways from the ways they in fact are. Does this imply that the same things exist in many possible worlds? Is there a problem about the identification of individuals across possible worlds, and if so what is it? There are different theoretical accounts of the relations between the individuals that exist in different possible worlds, and of the relations between particular things and the properties and relations that they exemplify.

1. Modal realism vs. actualism. The basic contrast between possibilist, or modal realist accounts of modality on the one hand and actualist accounts on the other is central to many of the more specific issues in the metaphysics of modality. According to the modal realist, there are literally many universes, individuated by the spatial and temporal relations between things in them. Two things count as worldmates – denizens of the same possible world – if and only if they are spatio-temporal relations between them. But for the actualist, everything that is real is actually real. Possible worlds are possible ways that a world might have been. The difference between the two kinds of theory comes out in the contrasting answers that they give to the following general challenge to the coherence of the idea of a merely possible world:

A merely possible world is a world that is not actual, which is to say a world that does not exist. But the possible worlds analysis of modality is committed to the existence of merely possible worlds, which seems to mean that it is committed to the existence of things that do not exist.

Any response to this challenge that seeks to defend the coherence of the account must distinguish a sense in which merely possible worlds exist from a sense in which they do not, and there are two very different strategies for making this distinction. The modal realist answers the question by distinguishing two different ranges for the quantifier – one unrestricted and one restricted. When we talk about absolutely everything that exists, we include a plurality of possible worlds (as well as merely possible donkeys, people and things). But we most often use the quantifiers so that they range over a restricted domain: "everyone" might, for example, mean all the people invited to the party. Even when we are making very general claims, we are often (according to the modal realist response to this challenge) restricting our quantifiers to things in our vicinity, broadly construed. Our vicinity, on this construal, includes that part of reality that is spatio-temporally connected with us. In this broad but restricted sense, there is only one possible universe that exists: the one we are in. But the other universes, like the actual people who were not invited to the party, are equally real.

For the actualist, the distinction is of a different kind. According to this theory, the only things that exists, in the most absolute and unrestricted sense, are actual things. The relevant distinction is not in the range of the quantifier, but in the kind of thing that one is talking about. Possible worlds, properly construed as things that there are many of, are more accurately labeled "possible states of the world", and states are the kind of thing that may be instantiated or exemplified. (For a given state, there may or may not be something that is in that state). We might distinguish a notion of "possible world" meaning a thing that exemplifies a given possible state of the world from a notion of "possible world" as the state itself – something that is perhaps not exemplified. Using "possible world" in the first sense, there is only one of them (within the domain of absolutely everything), while using it in

the second, there are (in this same domain) many, only one of which is exemplified.

The modal realist doctrine can be separated into two theses, one metaphysical and one semantic. The metaphysical thesis is that there is a rich plurality of spatio-temporally disconnected universes, rich enough to obey certain principles of recombination. (Roughly, for any two things in different universes, there will be a universe that contains intrinsic duplicates of both.) The semantic thesis is that statements about what is necessary and possible are properly interpreted by quantifiers that range over these universes (A sentence of the form "Possibly *P*" is true if and only if *P* is true in one of these universes.) Judged separately, both theses seem highly implausible. What reason do we have to believe in this extravagant ontology? And even if we did, what does it have to do with what is necessary and possible in the actual world? But while Lewis granted the prima facie implausibility of his doctrine (he took what he called "the incredulous stare" to be the most serious challenge to his metaphysical view), he argued that the two parts of the doctrine must be judged together, and that together they provide an indispensable foundation for a rich family of modal concepts. Despite its initial implausibility, the fruitfulness of the doctrine that provides this foundation is sufficient reason to believe that it is true.

Crucial to this defense of modal realism is the thesis that the rich family of modal concepts cannot rest on a more modest foundation. To use Lewis's rhetoric, the claim is that we cannot have the "paradise" that this family of concepts brings "on the cheap." In this context, Lewis criticizes several versions of the actualist alternative to modal realism, arguing that none of them is up to the job. All of the actualist accounts that Lewis considers assume that "possible worlds" must be *representations* of a world: either linguistic representations, something like scale models, or perhaps just simple and primitive representations. I think Lewis is right that a notion of possible world as representation cannot provide an adequate foundation for our modal concepts, but there are other alternatives that Lewis does not consider. Possible states of the world, or ways a world might be, are not representations of a world, but properties that a world might have. While properties allow for the distinction between existing and being exemplified that the actualist needs to distinguish the sense in which merely possible worlds exist from the sense in which they do not, properties differ crucially from representations in the following way: representations, whether pictorial, linguistic, mental, or of some other form, face a problem of INTENTIONALITY; it makes sense to ask, of a representation, what is it that explains why the representation has the representational CONTENT that it has? There is no analogous question about properties. One cannot intelligibly ask, of a property, what makes it *that* particular property, rather than some other one? I think Lewis's critique trades on the fact that the actualists do not have an answer to a question like this about the things they are calling "possible worlds".

In the context of Lewis's overall metaphysical picture, the thesis that possible states of the world are a kind of *property* does not provide an alternative to modal realism, since on Lewis's account, properties are classes, individuated by their extensions, and so a total way a world might be will be a unit set, with the world that is that way as its member (*see* CLASS, COLLECTION, SET). On this account of properties, there will be many possible total states of the world only if there are many things that are in those states. It is an irony of Lewis's modal realism that the metaphysically extravagant doctrine is grounded in Quinean ontological austerity – a rejection of any notion of property or attribute that cannot be identified with its extension. The actualist is committed to a more robust notion of property, and so needs an explanation of what properties are.

2. Reduction. Possible worlds, construed as concrete universes, are the fundamental primitive elements of the modal realist theory, and are clearly prior, in the order of explanation, to propositions, which are identified with sets of possible worlds (*see* CONCRETE/ABSTRACT). Necessity and possibility and the other modal notions are all

definable in terms of the properties of and relations between propositions. Is Lewis right to claim that this theory provides a reductive account of modality – an explanation of the modal in terms of the non-modal? This is a delicate question, since it is debatable what basic concepts count as modal, but I think Lewis's claim is a reasonable one. The reason is that the metaphysical component of the theory (the hypothesis of a plurality of parallel universes) is intelligible independently of the semantic analysis of modal concepts in terms of it. The parallel universes are individuated by spatio-temporal relations, and if it is fair to claim that the notion of a spatio-temporal relation is a non-modal notion, than he theory seems to offer a metaphysical characterization of the structure of reality in terms of concepts that modal skeptics should be willing to accept (even if they reject the substantive metaphysical claims made with those concepts). So I would concede Lewis's claim that he offers a reduction, but maintain that it is debatable whether this is a cost or a benefit of the overall account. It is not just the metaphysical commitments of the theory that elicit the incredulous stare; the semantic analysis of modal notions in terms of it also seems implausible, since it defines modal concepts in terms of things that, intuitively, seem to have nothing to do with modality, even if one were to accept the metaphysics. The intuitive resistance to the semantic component of the doctrine may derive from the judgment that modal notions are fundamental, and not properly reduced to something more basic. Compare the way one might react to a project of giving a reductive analysis of truth to something more basic (warranted assertability, perhaps, or what will be believed at the end of inquiry). Even if such a story could be spelled out in noncircular terms, one might judge that the analysis mistakenly categorizes a substantive claim as a definition. Even if it were correct that, at the end of inquiry all and only truths would be believed, this would not give us an account of what truth is (*see* THEORIES OF TRUTH). Similarly, I think it is reasonable to think that even if a principle of plenitude were true, so that everything that might happen does happen, somewhere and sometime, perhaps in a parallel universe, it would still be wrong to say that this is what possibility consists in.

I also agree with Lewis that no actualist attempt to explain possible worlds in non-modal terms (for example, as linguistic representations) can succeed. But most versions of modal actualism are not attempts to explain the modal in terms of the non-modal, since the basic notions of this kind of theory – whether they are propositions or total ways a world might be – are characterized in terms that presuppose modal notions. In fact, I think the notion of a *property*, which is used to say what kind of thing a possible state of the world is, is itself a modal notion: one grasps what property one is talking about to the extent that one has a sense for what it would be for that property to be exemplified, which is to understand a certain possibility.

3. Merely possible things. It seems at least prima facie reasonable to believe that there might have existed things that do not in fact exist. For example, Saul Kripke might have had seven sons, and if he had, then seven people who do not in fact exist would have existed (assuming that Saul Kripke actually has no sons). In the possible worlds framework, the general thesis is modeled by the claim that the domains of some possible worlds contain individuals that are not in the domain of the actual world. The modal realist has no problem with this thesis, since the actual world is just one place among others. Non-actual things are just things that are located in one of the other places. But for the actualist, the domain of the actual world includes everything that exists at all, in any sense, so it seems that actualism is at least prima facie committed to the thesis that everything that might exist does exist. This is the most serious challenge to the actualist conception. I will describe four strategies that different actualists use to respond to it:

The first actualist response begins by noting that to understand talk of possible individuals, we need a distinction that parallels the distinction between two senses of the term "possible world": just as actualists

must distinguish a way a world might be from a world that is that way, so they must distinguish individuating properties that an individual might have from the individuals that has those properties. While actualists are committed to the thesis that there are no *things* that might exist, but do not, they can allow that there are (and necessarily are) *properties* that are necessary and sufficient to determine a unique individual, but that are in fact uninstantiated. More precisely, the view is that there are properties X that meet the following condition: it is necessary that if there exists something that instantiates X, then that thing is necessarily identical to anything that instantiates X. The domains of the different possible worlds are to be understood, according to this response, not literally as domains of individuals, but as domains of properties of this kind – individual essences, or *haecceities* (*see* HAECCITY). Alvin Plantinga, who develops and defends this response, calls the domains "essential domains" (Plantinga, 2003). The basic structure of the orthodox Kripke semantics for quantified modal logic, with variable domains, is unchanged by this move; the difference is in the interpretation of the formal models. This response to the problem is simple, formally conservative, and successful, on its own terms, in reconciling actualism with intuitions about what might have been true. But it requires what some regard as a metaphysical extravagance: a belief in a special kind of property that carries with it the particularity of an individual, but that is also conceptually separable from the individual. We may have no problem understanding the property of being identical to Socrates, but one might reasonably think that this is an object-dependent property – a property that would exist only if Socrates did. But the haecceitist response to the problem holds that while we use the person Socrates to fix the reference of the property of being identical to Socrates (or a property that is necessarily equivalent to it), the property itself would exist even if he did not. And furthermore, there actually exist, on this account, properties of this kind that would be instantiated by Saul Kripke's seven sons in the possible worlds in which he had seven sons – properties whose reference could be fixed, in such a world, with a predicate of the form "being identical to *this* individual", where "*this* individual" refers to one of the seven sons. Plantinga grants that we may not have the resources, even in principle, to refer to particular uninstantiated haecceities, but we can talk about them in general terms, and that, he argues, is good enough.

The second response (defended in WILLIAMSON, 2002 and in Linsky and Zalta, 1994) is to reject the intuition that gives rise to the problem – that there might have existed things that do not in fact exist (as well as the intuition that there are some things that exist only contingently). This response avoids the problem, and as its defenders emphasize, it also allows for a modal logic that is much simpler than what is required when the domains vary from possible world to possible world. But of course it takes on the burden of explaining the divergence between the theory and conflicting intuitions about modal truths that seem compelling. How can it be made plausible that apparently temporary and contingent beings such as ourselves exist eternally and necessarily? How can we accept that there actually are things that might have been Saul Kripke's seven sons? The defenders of this strategy respond to the challenge by acknowledging that people and ordinary PHYSICAL OBJECTS are only temporarily and contingently concrete things, with a spatio-temporal location. In possible worlds and at times when one is inclined to say that the people and things do not exist, we should instead say that they exist, but lack the features that we are inclined to think are essential to being a person or a physical object. They are in no place, at those worlds and times, and are neither concrete things, nor abstract objects, but particular things that have the potentiality to be concrete things. This may seem a gratuitously extravagant metaphysics, but Williamson argues that it is entailed by principles that it is difficult to reject. The most controversial of the premises of Williamson's argument is the thesis that singular propositions (and identity properties, such as "being identical to

Socrates") depend for their existence on the things they are about. As we have seen, Plantinga rejects this thesis (which he labels "existentialism"), but I suggested that this is a serious cost of his account. But as Williamson shows, if we accept it, and also accept that a proposition is true only if it exists, then we must conclude that the singular proposition that Socrates does not exist could not be true, and this seems to imply that Socrates must exist.

One might try to avoid the uncomfortable choice between Plantinga's haecceities and Williamson's objects of pure potentiality by rejecting a presupposition that both positions apparently share. The need either for primitive individual essences or for object dependent propositions would be avoided if we were able to reduce individuals to their properties. So a third actualist response to our problem is to adopt some kind of BUNDLE THEORY of individuals. If this kind of account of individuals were defensible, then we could characterize possibilia in terms of ordinary universal properties and relations, rather than in terms of primitive haecceities, and no propositions would be dependent on particular individuals. But this kind of metaphysical doctrine has a problem accounting for the potentialities and counterfactual properties of particular individuals; we will say more about this problem below.

There is a fourth response that accepts the irreducibility of individuals to their properties and relations, and the object-dependence of singular propositions. It takes at face value the intuition that there might have been things other than those there are, and it avoids a commitment to individual essences. I think this is the best response to the problem, though it has its own counterintuitive consequences. The problems for this strategy come from an immediate consequence of the combination of the object-dependence of singular propositions with the contingent existence of individuals: that some propositions themselves are things that exist only contingently. If possible worlds are identified with maximal propositions, or maximal sets of propositions, then possible worlds themselves will be contingent objects. Propositions that are maximal in the sense that they entail every (actual) proposition or its contradictory may fail to be maximal in another sense: they may entail existential propositions without entailing any singular propositions that witness the existential claim. That is, this response claims that there may be cases where an existential proposition (Such as the proposition that Saul Kripke had a seventh son) is possibly true even though there is no singular proposition (no proposition of the form "x is Saul Kripke's seventh son") that is possibly true. I think this is right, but making sense of it requires a more radical reinterpretation of the standard semantic models than do the theories of Plantinga or of Williamson, Zalta and Linsky. And this response must accept the consequence that there are propositions (Such as the proposition that Socrates never existed) which are true with respect to some possible worlds in which the proposition itself does not exist. (Different versions of the fourth response have been defended in FINE (2005), and Adams (1981).

4. Modal properties. Even if we ignore merely possible individuals, there are problems with attributions of modal properties to actual individuals. *De re* modal claims, claims about what could or could not have been true of some particular things seem especially problematic since it is not clear how they could be true by convention, or by virtue of the relation of ideas. The possible worlds picture seems to offer a straightforward paraphrase of such claims: to say that David Lewis might have been a plumber, but could not have been a fried egg, is to say that there is a possible world in which David Lewis was a plumber, but no possible world in which he was a fried egg. But is the plumber in the other possible world really the same person as our own David Lewis? What is it about him explains his metaphysical incapacity to be a fried egg? Modal realists and actualists answer these questions in different ways.

If possible worlds are ways the world might have been then there is no implausibility in accepting the straightforward assumption that Lewis himself inhabits other worlds, since this is only to say that among the ways the world might have been but was not

are ways that David Lewis might have been. Kripke (1980) attempts to demystify counterfactual suppositions about particular individuals, arguing that nothing prevents us from simply stipulating, in specifying the counterfactual situation we are talking about, that it is a situation in which David Lewis is a plumber. But Kripke acknowledged that we might also specify a counterfactual situation in a way that does not explicitly identify a particular individual – in terms of the qualitative characteristics, origin, or constitutive parts of the individual, and that in such a case, we might then ask whether the individual we have specified is or might be some particular actual individual. It remains puzzling exactly what determines the answers to such questions.

If possible worlds are understood as other places, as the modal realist understands them, then it is no longer plausible to think that the inhabitants of the actual world will also be found in other possible worlds. The Lewisian modal realist explains modal properties of individuals – their capacities and dispositions, essential and accidental characteristics – in terms of the existence, in other possible worlds, of *counterparts* of the individual – individuals in other possible worlds who are similar, in relevant respects, to the given individual. According to counterpart theory, David Lewis himself existed only in the actual world, but he might have been a plumber in virtue of the fact that there is a possible worlds in which a person who is like him in certain specific respects was a plumber.

Actualists may also use counterpart theory, but for them there is no conflict between the counterpart analysis and the thesis that the plumber in the other possible world really is our own David Lewis. An actualist counterpart theorist may say, as Alvin Plantinga does, that the domains of other possible worlds should be thought of, not as sets of individuals, but as some kind of property that would have been instantiated by an individual, were the state of the world to have been realized. For the haecceitist, the relevant properties are individual essences, but an anti-haecceitist actualist might take the relevant individuating properties to be bundles of qualities, and reduce individual essences to such bundles and counterpart relations between them. An actualist might also adopt a counterpart framework, with a primitive counterpart relation, for methodological reasons: the aim would be simply to provide a framework that is neutral on controversial theses about essential and accidental properties, a framework in which puzzles about identity across times and worlds can be formulated in a perspicuous way. (Actualist counterpart theory is discussed in some papers in Stalnaker, 2003.)

Whether one is an actualist or a modal realist, and however one explains the apparent possibility of things that do not in fact exist, and the relation between particular individuals and the properties and relations that they exemplify, and might exemplify, there will remain a general puzzle about the nature and source of modal truth. If necessity is true in all possible worlds, what explains why there are just the possible worlds that there are? Both actualists and modal realists resist the idea that we can explain modal facts as conventional or semantic facts: conventions may determine that our words express certain propositions, but the propositions themselves are necessary or contingent, independently of the words that are used to express them. But actualists and modal realists also agree that to express substantive propositions is to distinguish between the possibilities – to locate the actual world in the space of all possible worlds – and this seems to imply that it is not possible to give a substantive characterization of what is common to all possible worlds. We won't have a clear grasp of the concept of metaphysical possibility until we see a way to resolve this tension.

BIBLIOGRAPHY

Adams, R.: "Actualism and Thisness," *Synthese* 49 (1981), 3–41.

Adams, R.: "Theories of Actuality," *Noûs* 8 (1974), 211–31.

Divers, J.: *Possible Worlds* (London and New York: Routledge, 2002).

Fine, K.: *Modality and Tense: Philosophical Papers* (Oxford: Oxford University Press, 2005).

Kripke, S.: *Naming and Necessity* (Cambridge, MA: Harvard University Press, 1980).
Kripke, S.: "Semantical Considerations on Modal Logic," *Acta Philosophial Fennica* 15 (1963), 83–94.
Lewis, D.: *On the Plurality of Worlds* (Oxford: Blackwell, 1986).
Linsky, B. and Zalta, E.: "In Defense of the Simplest Quantified Modal Logic," *Philosophical Perspectives* 8: *Logic and Language*, ed. J. Tomberlin (Atascadero, CA: Ridgeview, 1994), 431–58.
Plantinga, A.: *Essays in the Metaphysics of Modality* (Oxford: Oxford University Press, 2003).
Stalnaker, R.: *Ways a World Might Be: Metaphysical and Anti-metaphysical Essays* (Oxford: Oxford University Press, 2003).
Williamson, T.: "Necessary Existents," in A. O'Hear, *Logic, Thought, and Language* (Cambridge: Cambridge University Press, 2002), 233–51.

ROBERT STALNAKER

Persistence

Introduction

Things change. This much looks like a metaphysical and observational datum. By the proposition that things change we typically mean that things *survive* change – not all changes, but most. In other words, we live in a world in which there is both change and sameness. My car was red; I have given it a coat of green paint; now it is green. One of the car's qualities has changed, and to that extent the car itself has changed. But we would all accept that it is still the same car. The standard way of putting this philosophically – though not a way we often describe it – is to say that the car has *persisted* through a change, in this case of color.

Yet it is not just change that compels metaphysicians to wonder about persistence. It may be that, as ARISTOTLE held, time is the measure of change, and so without change there could be no time and hence no persistence – since persistence occurs *in* time (extra-temporal existence, such as GOD's, would not on this view be a kind of persistence) – yet things persist even when undergoing no macroscopic change such as that of color. What is it for a material object to persist pure and simple? Is this a misconceived question because persistence is too basic a phenomenon to yield to analysis? Or can the metaphysician say something informative about what it involves? Is there something inherently strange, or even paradoxical, about the concept of persistence such that we ought to deny that anything really persists? Or can we retain the idea of persistence and instead deny that anything changes since it is change, rather than persistence itself, that raises insoluble problems?

Spatio-temporal continuity

The standard approach to analyzing persistence is in terms of spatio-temporal continuity (Coburn, 1971; Swinburne, 1968/1981). The idea is that an object *F* persists through a temporal interval if and only if it traces a spatio-temporally continuous path through that interval. Tracing a spatio-temporally continuous path is then defined in terms of overlap between every pair of adjacent spatio-temporal regions enclosing *F* during the interval. The approach is intuitively plausible inasmuch as we do tend to think of persistence in terms of some kind of continuity, or perhaps continuous history, involving the persisting object. We tend to associate diachronic distinctness (distinctness over time), not just synchronic distinctness (distinctness at a time), with breaks in continuity, for example between my car and my house: there is no single continuous path traced by both of them; their spatio-temporal histories are discontinuous.

It turns out, however, that it is far more difficult to spell out an adequate continuity criterion of identity over time than it seems, since it has to rule out obvious counter-examples. For instance, take a single-celled organism such as an amoeba, which reproduces by binary fission. It produces daughter organisms neither of which are, it seems, identical to the original; yet one can trace a continuous path between the pre-fission amoeba and each of its descendants. Or consider a marine flatworm, cut into two

segments that then grow into new worms. Perhaps the continuity in these cases is too weak, since we could specify a strong form of continuity such that the difference in overlap between any adjacent regions was indefinitely small, which does not seem to obtain in an instantaneous division of the sort just mentioned. But strong continuity also has counterexamples both to necessity and sufficiency for diachronic identity. As to necessity, consider an instantaneous loss or gain of parts: a tree has a branch lopped off, yet it still persists, though the continuity is only weak. As to sufficiency, consider the infinite series beginning with a tree, all of the other members of which are decreasingly smaller parts of the tree – the lump of plant matter minus a millimetre of wood, the lump minus two millimetres, and so on, where the series fully constituted is a real continuum of spatio-temporal parts measured along some dimension, such as length or width. We can specify a strongly continuous path, but we do not want to say that the tree, although strongly continuous with its parts, is identical to any of them.

The obvious move, at least to counter whole–part tracing confusions, is to place some sort of sortal restriction on what must be in the path: since parts of trees are not trees, a series such as that just given would not constitute a single persisting object. But there are other examples that make difficulty even here (SHOEMAKER, 1979; Forbes, 1985, pp. 152–9; the examples go back in some respects to KRIPKE's unpublished lectures on identity over time from 1978). Consider a homogeneous rotating sphere. Take one of its segments, i.e., one of its physical parts that moves with the sphere. This is clearly a single, persisting object, namely a sphere segment. Now imagine a light that constantly illuminates one single region with the same surface area as the segment. The segment passes through the illuminated region one time for every complete rotation. But during the period of one rotation, an infinite series of distinct sphere segments also pass through that single, illuminated region. They occupy a strongly continuous path, and they all fall under the same sortal *sphere segment*, yet they are not a single, persisting sphere segment. Such cases are easily multiplied and provide a formidable challenge to continuity theories of identity. Perhaps the causal relations and counterfactual dependence between the segments is relevantly different from those between the single segment at one time in its history and at another, but spelling this out is no easy matter.

Ought we to take a different approach to analyzing persistence? We might take a cue from BUTLER's famous criticism of the memory criterion of personal identity over time (Butler, 1975, p. 100), that "consciousness of personal identity presupposes, and therefore cannot constitute, personal identity". We might argue (Oderberg, 1993; Merricks, 1998) that spatio-temporal continuity gives us *evidence* of persistence but does not *constitute* it. For continuity always presupposes identity, inasmuch as the objects related by continuity (the car at t_1, the car at t_2) are themselves persisting objects – so how can continuity be used to *analyze* persistence if persistence is always part of what is described in describing a case of continuity? If continuity can only have evidential force – being a *symptom* of persistence but not a *criterion*, as it is sometimes put – then the evidence will be defeasible, and in some cases easily so. Where it is absent, moreover, we may still have good grounds for believing identity to obtain: imagine the radical disassembly and reassembly of an object, or its vanishing and reappearing. (Is the latter a metaphysical impossibility? It is certainly conceivable.) Absent any other viable analyses of persistence, we might take it to be a brute fact, an unanalyzable phenomenon. We usually know it when we see it, though we do make mistakes of reidentification.

Temporal parts

A defender of continuity, however, will not be content with the circularity objection to the proposed analysis. We do not have to think of the relata of the continuity relation as themselves persisting objects: they are, rather, *temporal parts* of persisting objects, terminating ultimately in *instantaneous*

temporal parts. Each continuous path is occupied by temporal parts, and each temporal part is itself analyzable in terms of parts of shorter duration, also on a continuous path. The circularity is only apparent, since the termination of the analysis lies in parts that do not themselves persist – we might think of these as *points* in space and time, occupied by certain qualities.

Temporal part theory, often known as four-dimensionalism or (misusing an old word) the theory of *perdurance*, has many defenders. (*See*, for a small sample: QUINE, 1950; LEWIS, 1986; Forbes, 1987; Heller, 1990; ARMSTRONG, 1997.) The general idea is that just as persisting objects have spatial parts (the wheels on my car, the branches of the tree) so they also have temporal parts (that part or stage of my car from t_1 to t_2, the part or stage of the tree from t_3 to t_4). The temporal parts are usually designated by a kind of hyphenated singular term: "my car-from-Monday-to-Wednesday", "the tree-from-Thursday-to-Saturday", and so on for any persistent and for any times however specified.

The idea has intuitive appeal, since we know that objects have spatial parts, and space and time are in many ways similar. Moreover, it seems that contemporary space–time physics, with the theory of relativity at its core, is at least congenial to temporal parts if not committed to them. There is, it might be claimed, no space and no time – only space–time. Space–time is an ontological unity, with objects spread out across both the three spatial and one temporal dimensions, all of which are features of a single "block", with objects being (on the favorite metaphor) something like "worms" stretched out across the block, divisible into "segments". These segments, speaking accurately, are supposed to be spatio-temporal parts: there are no *purely* spatial parts, and no *purely* temporal parts, but these spatio-temporal parts are what are *called* temporal parts on the four-dimensionalist way of looking at the universe.

How can the four-dimensionalist get persistence from temporal parts? He might simply say that a series of temporal parts constitutes a persisting thing if and only if it tracks our best intuitive, pre-theoretical judgments about what persisting things there are; in other words, temporal part theory should leave our reidentification practices undisturbed. Ordinary persistence aside, moreover, if we think that a certain object might, say, vanish and reappear after an interval, we could count the series of its temporal parts before disappearance and after reappearance as constituting a single persistent. This view of persistence could be supplemented, or to some extent modified, by a mixture of ontology and evolutionary theory along these lines: every materially occupied portion of space–time, no matter how heterogeneous, constitutes an object. We humans, for survival purposes, "gerrymander" certain portions of space–time and call these particular, privileged "worms" the persisting objects. (*See*, for example, Quine, 1981.)

The debate about the existence of temporal parts and their putative explanatory role as regards persistence continues unabated. The intuitive appeal of four-dimensionalism has captured the imagination of many metaphysicians, but it has to face some serious objections. (For some of the critics, *see* Geach, 1972; CHISHOLM, 1976: Appendix A; THOMSON, 1983; Oderberg, 1993; LOWE, 1999, pp. 114–18.) For instance, the thought that just because an object has spatial parts so it must have temporal parts is specious. For in order to generate a sufficiently convincing analogy between space and time to motivate the thought, it turns out that one has to *presuppose* the existence of temporal parts in the first place (as can be seen in Taylor, 1955; discussed in Oderberg, 1993, pp. 97–103; *see also* Meiland, 1966). Second, is it true that space–time physics commits us to temporal parts? To say that Minkowskian space–time geometry has shown, as Minkowski himself thought, that space and time are "mere shadows" of an underlying unified space–time (Minkowski, 1952, p. 75) could be seen as a metaphysical step too far since the spatial and temporal dimensions are given differing mathematical treatments in relativity theory. One should, in addition, be careful about drawing metaphysical conclusions from

physicists" use of terms such as "world-line", "space–time worm", and the like, since where these terms appear in their work persistence is nearly always *presupposed* as a more fundamental concept rather than explained or *analyzed* in those terms. It is in fact very difficult to motivate a metaphysic of temporal parts from space–time physics (Rea, 1998).

It might further be argued, independently of considerations from the physics of space–time, that the very concept of a temporal part of a persisting object is of dubious coherence. To be sure, the temporal part skeptic does not deny that *some* things have temporal parts: events paradigmatically have them (the first half of the battle, the last five minutes of the opera), as do processes (the first hour of a compound's dissolution in water) and histories (the medieval history of Portugal; the first half of my life). What the skeptic denies is that *persisting objects* have such parts, and while events and processes involve objects that persist, they themselves do not persist. So what sense can be made of the very idea that a persisting object could have temporal parts? The hyphenated singular terms mentioned above are a philosophers' invention; not all such inventions are bad, of course, but we should not infer from their existence that what they purport to refer to exists as well. For how could we – how even could God – distinguish between a putative temporal part of an object and a temporal part of that object's history (or career, as it is sometimes called) with exactly the same temporal boundaries?

The critic needs to be more precise, though. The history of my car from Monday to Tuesday involves more than just the car itself: there are all of its relations to other objects that need to be included in that history. The putative temporal part of the car itself from Monday to Tuesday, however, is supposed to involve only what is within the car's spatial boundaries. The critic, however, can reply as follows. Call that part (not temporal, not exactly spatial – let's think of it as quasi-spatial) of my car's Monday-through-Tuesday history that involves only the car itself and its intrinsic features its *intrinsic history*. Hence we factor out, for instance, that part of its history involving its being parked by the kerbside or its being owned by me; we include its being green, curved on top, and having five windows. Now this intrinsic history is a genuine part of the car's total history. If you wanted to, and you knew my car well enough, you could write a rather boring narrative of its history from dawn on Monday until dusk on Tuesday. Now, says the skeptic, what is to distinguish *this* history of the car from Monday to Tuesday from the supposed car-from-Monday-to-Tuesday? Could God, let alone we, tell them apart? But, comes the reply, the temporal part of the *car* is a physical object, whereas the temporal part of its *history* is not. It is the physical temporal part that makes the history true. If the car's history involves its being green on Monday and receiving a coat of red paint on Tuesday, the Monday-Tuesday temporal part will have a green sub-part existing on Monday and a red sub-part existing on Tuesday. Yet the skeptic will insist that there just is no ontological *room* for such objects. What makes the car's history what it is from Monday to Tuesday is just the car itself and what is true or false of it: it is green on Monday, red on Tuesday. This is what makes it the case that it has a history with the following temporal parts – the Monday history, in which it is green, and the Tuesday history, in which it is red (and, of course, the temporal part overlapping these in which it is changed from red to green). What room is there for temporal parts *of the car itself*?

Among various other objections, a couple more are worth raising. Remember that to avoid circularity in the analysis of persistence, temporal parts will have to terminate in instantaneous entities of which all the rest (those with duration) are composed. Yet what sense can be made of an instantaneous temporal part? Calling it a space–time point (with or without qualities "associated" with or "true" of it) does not clarify matters. If instantaneous stages are really that – durationless – then how can they constitute an object with duration any more than dimensionless points can constitute a region with dimension? Wasn't ZENO right all along? One reply is to appeal to the Aristotelian notion

of potential infinity: the instantaneous stages are no more than *limits* of a process of potential division, but it is not as though such things have any actuality. The notion might be a good one, but can the four-dimensionalist appeal to it given that he has non-circularly to analyze persistence in terms of stages? No one who accepts the distinction between actual and potential infinity would want to *analyze* a line in terms of dimensionless points. But the friend of temporal parts needs just such an analysis if he is to avoid being left with an unanalyzed remainder of persisting, i.e., non-instantaneous, temporal parts.

Another point concerns whether four-dimensionalism denies the phenomenon it seeks to explain. If persistents are just sums of stages, does anything really persist in the first place? Rather than genuine persistence, doesn't the friend of temporal parts offer us no more than a series of creations and annihilations, with new matter literally springing into existence ex nihilo all the time (Thomson, 1983)? The temporal parts theorist might bite the bullet here, taking his account to be eliminative; though he would be committed to implausible claims about creation and annihilation. Or he might say that this interpretation is true only if he is a *presentist* about time, according to which only the present moment and what happens in it are real. More congenial to his position, though, is *eternalism*, according to which all moments of time and what happens in them are equally real. Matter, on the latter view, does not keep vanishing and springing into existence; rather, the sum of stages making up a persistent is, as it were, given "all at once" – not simultaneously, but with equal reality. There is no temporal becoming: the space-time worm just exists with its spatio-temporal dimensions. Wherever one of the segments is in space-time, so is the persisting object present, just as my car is present wherever one of its spatial parts is. The skeptic still worries that an eternalist view also denies persistence: there is no persistence where there is just creation and annihilation, but equally no persistence where the object is viewed simply as a block in space–time. Moreover, on both presentist and eternalist interpretations, even if persistence or something approximating it is maintained, can the equally basic phenomenon of *change* can be accounted for? We will return to this shortly. First, let us briefly consider a couple of other accounts that can be given of persistence.

Stage theory

A view quite similar to standard four-dimensionalism/temporal parts theory/perdurance is sometimes called "stage theory" or "exdurance" (Haslanger, 2003). According to this theory (Sider, 2001; Hawley, 2001), there are indeed temporal parts of things other than events, processes, and histories, and there are space-time worms consisting of series of such parts. The basic four-dimensional framework is accepted. What the stage theorist denies, though, is that any of these worms are identical to what we identify as ordinary persisting things. When we talk about persistents we are not talking about worms but about the temporal parts themselves. What we think of as persistents are no more than stages.

Of the various motivations for this approach, an important one is to avoid what is seen as a problem about spatio-temporal coincidence. So, assuming (perhaps rashly) that personal fission is possible via a split-brain operation and transplant, suppose that I undergo this procedure and two people each get half of my brain. When they awake, each is psychologically continuous with me. Call the new persons Bill and Ben. Bill is to be tortured, Ben is to live in pleasure. Should I be worried about what will happen to me after the operation, or not concerned, or both, or neither? On the standard four-dimensionalist model, the most likely interpretation of events is that two worms exist before, during, and after the operation – the one including the temporal parts of Bill post-fission and me pre-fission, these being connected by psychological continuity; and the one including the continuous stages of me and Ben. But this means that pre-fission, there are two space–time worms overlapping – the me-Bill worm, and the me-Ben worm. But if persistents are worms, and persistents include persons, then

it looks like there are two persons overlapping – coinciding in space and time – before the fission occurs. How, though, can two persons be in the very same place at the same time? And what thoughts am I likely to have – a single ambiguous thought about future pain and future pleasure for me, or two thoughts? Wouldn't the pronoun "I" be ambiguous pre-fission? Could two persons share a single thought, ambiguous or not? Surely I wouldn't notice any ambiguity when thinking about my fate.

This interpretation of fission has too many problems, according to the stage theorist. What we need to say is that there is a single person pre-fission, and that person is a stage, and that stage will be both continuous with Bill-stages and continuous with Ben-stages. In this sense it is true to say, for the stage theorist, that I will be Bill and I will be Ben, but there is only one of me prior to fission. To say that I will be Bill (and will be Ben) is akin to what a counterpart theorist such as Lewis says about modal statements. The counterpart theorist interprets a statement such as "I could have been smarter" as meaning that in some possible world there is a counterpart of me who is smarter than I am (in the actual world). Similarly, for the stage theorist to say that I will be Bill is to mean, properly interpreted, that the stage that I am is continuous with some future Bill-stage. Since the relation is duplicable, it can hold simultaneously of the stage that I am and both future Bill-stages and future Ben-stages.

To critics, the stage view fares little better than standard four-dimensionalism. Pre-fission (at t), I can truly say that after fission (at t_1) I will be Bill and that after fission I will be Ben (but not, on a necessary revision of standard reasoning about tense, that I will at t_1 be Bill and Ben: *see* Sider 2001, pp. 201–2). Yet if I will be Bill at t_1, is it not the case that there must at t_1 *be* a stage that *is* me, and that that stage is a Bill-stage? Yet the same is true for myself and Ben: if I will be him, then at t_1 there must be a Ben-stage that *is* me. Yet I cannot be both of them since they are distinct! And by the hypothesis of equal continuity there is no good reason to say that I am either. All the stage theorist allows is that at t_1 there is Bill, who *was* me, and Ben, who *was* me, and these relations are explained in terms of the duplicable relation of continuity. It is hard to see how genuine identity – and hence persistence in anything like a recognizable form – gets into the picture. If, before fission, I really *am* a certain stage, then how, without violating the necessity of identity, can it be the case that there is any stage *after* fission that *is* me if the pre-fission stage that I am now no longer exists then? So isn't it the case, on the stage view (assuming necessity of identity) that I simply cease to exist at fission? In which case I will *not* be Bill and I will *not* be Ben, in any recognizable sense of those propositions.

The stage theorist will accept that what he posits is not genuine identity; as one such theorist puts it, "claims of identity between things at different times make sense, even though they are false" (Hawley, 2001, p. 156). Yet the stage theorist wants to preserve the commonsense belief that persons and other persistents do exist. So on the theory, persons (for example) do exist but no identity statements about them are literally true! Well, it takes time to have a thought – even the simplest and most fleeting of thoughts – but it cannot then be literally true that it is I who has any thoughts. For persons are stages, and the only real stages are instantaneous ones. One can call a three-hour stage of me a stage, but it only has the title honorifically, or we might say derivatively. *It* literally has no thoughts, rather it is made up of instantaneous stages with mental properties suitably related in some unspecified (and arguably unspecifiable) way. But what possible mental properties could an instantaneous stage have, if even the briefest of thoughts takes time? Presumably, moreover, there is only one of me. But if I am a stage, *which* stage? It seems that stage theory retains all of the vices of standard four-dimensionalism but loses any virtues, for at least the standard worm theorist hold that there is precisely one of me, and that this single person is a four-dimensional sum of stages. On the stage view, it looks as though eliminativism about persons and other persistents is the unavoidable consequence: not

a happy result for a position that wishes to help itself to the commonsense belief that persons, cars, trees, and the other familiar objects of our universe do indeed exist.

Endurance

Theorists of persistence usually speak as well of "endurantism", taken as the view of virtually all those who deny that objects have (or are) temporal parts, so rejecting four-dimensionalism of any stripe. There are many such metaphysicians, but whether there is a theory around which they all rally is dubious. The believer in endurance rejects four-dimensionalism for the reasons already given and more. Stage theory, as we have seen, denies literal persistence. Standard worm theory takes there to be persistents – sums of continuous stages – but, according to the critic, it denies the reality of change (*see*, for example, Lombard, 1994, replying to Heller, 1992; *see also* Oderberg, 2004). All there is, on four-dimensionalism, is *replacement* of one temporal part by another, or *addition* of one temporal part to another – but neither replacement nor addition are genuine change. When my red car is painted green, a red car-stage (or series of such stages) is replaced or added to by a green car-stage (or series of such stages). As PEIRCE put it, "Phillip is drunk and Phillip is sober would be absurd, did not time make the Phillip of this morning another Phillip than the Phillip of last night" (Peirce, 1931, 1.494).

The endurance "theorist" wants to retain both the commonsense belief that there is literal persistence and the commonsense belief that there is genuine change throughout that persistence. In other words, one and the same object literally has a property and loses it. No four-dimensionalist theory can hold on to both of these beliefs. True, for the worm theorist my car does exist at every time at which any of its temporal parts do, but the properties it gains and loses are gained and lost *only* in virtue of there being distinct stages that have and do not have those properties respectively. My house also exists at every place at which it its spatial parts do, and many of its intrinsic properties are had only because one or more of its parts has those properties: it is warm because parts of it are warm, it is brick because it has parts made of brick, and so on. Now whether the worm theorist can account for all the properties of a thing in terms of properties of its temporal parts is highly questionable (*see* Zimmerman, 1998), but even restricting ourselves to those that she can so account for, why shouldn't we say that my house changes across space as well, since it has distinct spatial parts with different properties? My house does in one sense *vary* across space, but the endurantist holds that not all variation is change, and that something crucial is lost when change is defined (as it so often is in metaphysics texts) as the mere having of a property by an object at one time and its lacking it at another. For if that is all there is to change, then objects change across space as well, since why on this view should having a property at a time be significantly different from having it at a place? But change is a fundamentally *dynamic* phenomenon, involving a real *transition* of a thing itself from one state to another. Mere addition, replacement, and/or distinctness of parts do not capture this phenomenon.

The endurantist is often asserted to hold as a theoretical commitment that a persistent is "wholly" present at different times (Lewis, 1986). If there is any theory here, it is the denial of four-dimensionalism. But it is not the assertion of a recondite metaphysical state, merely the belief that one and the same object literally exists through time, itself having properties at some times that it loses at others. There are, though, theoretical consequences of this commonsense view. For example, endurance rules out any approach to fission cases that posits coinciding pre-fission objects, or violation of the standard logic of identity (the idea that I will be both Bill and Ben is a non-starter), or any relation weaker than strict identity as capturing "what matters" as between me and my post-fission descendants. Since identity cannot hold between myself and both Bill and Ben, the only option for the endurantist is to deny that I continue to

exist after fission: I am neither Bill nor Ben. Since the fission case has psychological continuity built into it, this means denying that I really will be psychologically continuous with both Bill and Ben and hence denying that genuine fission is possible, on the assumption that psychological continuity is present wherever personal identity is and vice versa. (The endurantist could hold the weaker position that psychological continuity is defeasible evidence of personal identity, and simply claim that in the case of myself, Bill, and Ben, the evidence is defeated.)

Change: metaphysics and semantics

What, then, of the phenomenon of change? In contemporary discussion, following Lewis (1986), theorists set up what is called the "problem of temporary intrinsics". The idea is that the following propositions are incompatible: (1) that objects (such as my car or me) persist through change; (2) that, across the same dimension of change, the intrinsic properties involved in an object's change are incompatible (being red and green, or red and non-red, round and square, round and non-round, etc.); (3) no object can possess incompatible properties. The problem is how, in analyzing change, all of these very plausible claims can be held true. KANT, for one, gave voice to a worry about how there could be a "combination of contradictorily opposed predicates in one and the same object" (Kemp Smith, 1933, A32/B48, p. 76).

A concern about the way in which this problem has been tackled is that two distinct issues have tended to be conflated: the *semantic* one of how to represent sentences describing change in such a way that they do not state a contradiction, and the *metaphysical* one of how change should be understood in such a way that no contradiction is assumed or implied. Semantics and metaphysics are not the same thing, and so the identity theorist needs to be careful to separate these issues. Taking the metaphysical one first, the obvious target is the third proposition. Does the Law of Non-contradiction state that nothing can have incompatible properties? Not at all. The locus classicus for the law, followed by virtually all philosophers ever since, is Aristotle, who affirms: "[T]he same attribute cannot *at the same time* belong and not belong to the same subject *and in the same respect*" (*Metaphysics*, Book Gamma, sect. 3, 1005b19; Ross, 1928; emphasis added). When my green car is painted red it certainly does not, at any one time, possess incompatible properties – the change itself ensures that the law is not violated, nor could it be. Hence it looks as though the "problem of temporary intrinsics" is spurious: why would anyone want to affirm the third proposition unless they had not thought carefully about the Law of Non-contradiction in the first place, or they wanted something to puzzle about for the sake of it?

Similarly, though less obviously, change does not involve any violation of Leibniz's Law. This law (more precisely that half of the law called the Indiscernibility of Identicals) states that if x and y are identical then they share all their properties. Some writers (e.g., Heller, 1992) argue that if an object such as my car is red at t and green at t_1, then my car at t is discernible from my car at t_1, the first being red and the second non-red. But this cannot be, so the properties must be possessed by numerically distinct temporal parts (united into a single four-dimensional worm). Yet Leibniz's Law is entailed by the Law of Non-contradiction: no object can both possess a property and lack it at the same time and in the same respect. So if x and y are identical, i.e., the same object, that object x (y) cannot be F and not-F at the same time and in the same respect. So it must be that if x is F, then it (y) is F at the same time and in the same respect, and vice versa. So if change violates Leibniz's Law, it violates the Law of Non-Contradiction, which it cannot do. For if change meant that x and y really did not share all their properties, though they were one and the same object, a contradiction would result. To say that Leibniz's Law still allows discernibility at *different* times, so is not entailed by The Law of Non-contradiction, is to miss the point. For if y has a property at t_1 incompatible with a property that *it* (x) has at t, there will be a contradiction if x does not also have the

property that y has at t_1; it just won't have it at t. In other words, Leibniz's Law does not mean that objects cannot *change*: if such change is just what we mean by "discernibility at different times", that is harmless. But if we intend something more by the expression, e.g. that the property y has at t_1 is a property that x (=y) *lacks* at t_1, even given that x exists at t, we end up in contradiction. So: my car at t is red at t, and my car at t is green at t_1; my car at t_1 is green at t_1, and my car at t_1 is red at t. My car does not *at any time* possess incompatible properties, though it does so at different times. But having incompatible properties at different times does *not* mean that x has any property that y lacks, and conversely.

This is where the semantic problem rears its head, though for all its interest we can only consider it briefly. The problem is how, semantically, to represent my car's change in such a way that no contradiction is stated or implied. What do these expressions such as "at t", "at t_1", and so on, mean? Does it matter where they are placed in a sentence stating property possession? There are at least three alternative proposals for dealing with this. The four-dimensionalist (including the stage theorist) applies the metaphysics to the semantics: the temporal qualifiers are affixed to the subject terms so as to block a contradiction. My car-at-t is red and my car-at-t_1 is green. The subject terms denote temporal parts of my car, so no one thing literally possesses incompatible properties at any time. For reasons already given, the temporal part skeptic will reject this approach. Another is called adverbialism (Johnston, 1987; Hanslanger, 1989): the temporal qualifiers are attached to the copula. Hence my car is-in-the-t-way red and my car is-in-the-t_1-way green. The same object possesses incompatible properties but in different ways, and these different ways of property possession remove the contradiction. It is difficult to get a grip on whether adverbialism has any metaphysical implications, and if so what they are. One semantic criticism is that the adverbialist has to give an account of temporal adverb dropping. Since we can usually drop adverbs and preserve truth (Fred runs fast, therefore Fred runs), why can't we drop temporal adverbs? But if we do so we end up with a contradiction again – my car is red and non-red – so must the adverbialist say the temporal adverbs cannot be dropped simply to avoid contradiction? If so, the justification for the solution looks circular: introduce temporal adverbs to block contradiction, but then exclude the standard semantic rule for dropping adverbs because otherwise there would be a contradiction. (*See* further Oderberg, 2004, and also Merricks, 1994 for criticism of adverbialism.)

A third approach is called "sententialism" (Oderberg, 2004; *see also* Myro, 1986, who uses sentential temporal operators but for a different purpose). Taking the separation of semantics from metaphysics seriously, the sententiaist holds that the temporal operators in sentences describing change are affixed to atemporal predications. At t, my car is red and at t_1, my car is non-red. The temporal operators create something like an opaque context: not strictly, since the context is still extensional (it doesn't matter what co-referring subject term I use for the sentence to be true), but the operators cannot be dropped. Why not? The main reason is that atemporal predications for changeable objects are incomplete – they do not state facts about the objects, only incomplete information that needs supplementation to make sense. So no inference can sensibly be made for such objects from a temporal to an atemporal predication. This is reinforced by the semantic fact that when we make predications that are not explicitly temporal – "My car is red" – there is always taken to be an *implicit* reference to the present, since otherwise the statement would be radically incomplete and truth unevaluable. Hence semantics is on the side of the sententialist, whereas the adverbialist has the standard rule in favor of adverb dropping to contend with.

See also the A–Z entries on BEING AND BECOMING; CHANGE; CONTINUANT; CONTINUITY; IDENTITY; PERSONS AND PERSONAL IDENTITY; SPACE AND TIME; TEMPORAL PARTS, STAGES.

BIBLIOGRAPHY

Armstrong, D.M.: *A World of States of Affairs* (Cambridge: Cambridge University Press, 1997).

Butler, J.: "Of Personal Identity," Appendix 1 of *The Analogy of Religion* (1736); repr. in *Personal Identity*, ed. J. Perry (Berkeley: University of California Press, 1975).

Chisholm, R.: *Person and Object* (La Salle, IL: Open Court, 1976).

Coburn, R.: "Identity and Spatiotemporal Continuity," in M. Munitz, ed., *Identity and Individuation* (New York: New York University Press, 1971), 51–101.

Forbes, G.: "Is There a Problem about Persistence?," *Proceedings of the Aristotelian Society*, suppl. vol. 61 (1987), 137–55; repr. in Haslanger and Kurtz (2006).

Forbes, G.: *The Metaphysics of Modality* (Oxford: Clarendon Press, 1985).

Geach, P.: "Some Problems about Time," in his *Logic Matters* (Oxford: Blackwell, 1972).

Haslanger, S.: "Endurance and Temporary Intrinsics," *Analysis* 49 (1989), 119–25.

Haslanger, S.: "Persistence through Time," in M.J. Loux and D.W. Zimmerman, ed., *The Oxford Handbook of Metaphysics* (Oxford: Oxford University Press, 2003), 315–54.

Haslanger, S. and Kurtz, R.M., ed.: *Persistence: Contemporary Readings* (Cambridge, MA: Bradford Books/MIT Press, 2006).

Hawley, K.: *How Things Persist* (Oxford: Clarendon Press, 2001).

Heller, M.: *The Ontology of Physical Objects* (Cambridge: Cambridge University Press, 1990).

Heller, M.: "Things Change," *Philosophy and Phenomenological Research* 52 (1992), 695–704.

Johnston, M.: "Is There a Problem about Persistence?," *Proceedings of the Aristotelian Society*, suppl. vol. 61 (1987), 107–35.

Kemp Smith, N. (trans.): *The Critique of Pure Reason*, by Immanuel Kant (London: Macmillan, 1933; originally published 1781 (A), 1787 (B)).

Lewis, D.: *On the Plurality of Worlds* (Oxford: Blackwell, 1986).

Lombard, L.: "The Doctrine of Temporal Parts and the 'No-Change' Objection," *Philosophy and Phenomenological Research* 54 (1994), 365–73.

Lowe, E.J.: *The Possibility of Metaphysics: Substance, Identity, and Time* (Oxford: Clarendon Press, 1999).

Meiland, J.W.: "Temporal Parts and Spatiotemporal Analogies," *American Philosophical Quarterly* 3 (1966), 64–70.

Merricks, T.: "Endurance and Indiscernibility," *The Journal of Philosophy* 91 (1994), 165–84.

Merricks, T.: "There Are No Criteria of Identity Over Time," *Noûs* 32 (1998), 106–24.

Minkowski, H.: "Space and Time," in H.A. Lorentz, A. Einstein, H. Minkowski, and H. Weyl, *The Principle of Relativity* (New York: Dover, 1952); address originally given in 1908.

Myro, G.: "Identity and Time," in R. Grandy and R. Warner, ed., *Philosophical Grounds of Rationality* (Oxford: Clarendon Press, 1986), 383–409.

Oderberg, D.S.: *The Metaphysics of Identity Over Time* (New York: St. Martin's Press, 1993).

Oderberg, D.S.: "Temporal Parts and the Possibility of Change," *Philosophy and Phenomenological Research* 69 (2004), 686–708.

Peirce, C.S.: *Collected Papers*, vol. I (Cambridge, MA: Harvard University Press, 1931).

Quine, W.V.: "Identity, Ostension, and Hypostasis," *The Journal of Philosophy* 47 (1950), 621–33; repr. in his *From a Logical Point of View* (New York: Harper and Row, 1961), and in Haslanger and Kurtz (2006).

Quine, W.V.: "Worlds Away," in his *Theories and Things* (Cambridge, MA: Harvard University Press, 1981).

Rea, M.C.: "Temporal Parts Unmotivated," *The Philosophical Review* 107 (1998), 225–60.

Ross, W.D. (trans. and ed.): Aristotle's *Metaphysics*, vol. VIII of *The Works of Aristotle Translated into English*, 2nd edn. (Oxford: Clarendon Press, 1928).

Shoemaker, S.: "Identity, Properties, and Causality," in P. French, T. Uehling, and H. Wettstein, ed., *Midwest Studies in*

Philosophy 4 (Minnesota: University of Minnesota Press, 1979), 321–42; also in his *Identity, Cause, and Mind* (New York: Oxford University Press, 2003).

Sider, T.: *Four-Dimensionalism: An Ontology of Persistence and Time* (New York: Oxford University Press, 2001).

Swinburne, R.: *Space and Time* (New York: St Martin's Press, 1968; 2nd edn., 1981).

Taylor, R.: "Spatial and Temporal Analogies and the Concept of Identity," *The Journal of Philosophy* 52 (1955), 599–612.

Thomson, J.J.: "Parthood and Identity Across Time," *The Journal of Philosophy* 80 (1983), 201–20; repr. in Haslanger and Kurtz (2006).

Zimmerman, D.W.: "Temporal Parts and Humean Supervenience: The Incompatibility of Two Humean Doctrines," *Australasian Journal of Philosophy* 76 (1998), 265–88.

DAVID S. ODERBERG

Realism and Antirealism about Abstract Entities

1. *What Is Realism?*

Realism about abstract entities, in its most general form, asserts – and antirealism denies – that there are such things. This simple formulation calls for some explanatory comment.

(1) Neither realism, nor its denial, need be an all or nothing affair. Realists have asserted, and their opponents have denied, the existence abstract entities of several different kinds – UNIVERSALS, mathematical entities such as numbers and sets, propositions, and various others (*see* NUMBER; CLASS, COLLECTION, SET; PROPOSITION, STATE OF AFFAIRS). Realists may be selective about the kinds of abstract entities whose existence they assert, and antirealists may likewise be selective – denying the existence of one kind of abstract entities, while remaining agnostic about, or even accepting, the existence of another. QUINE, for example, was a realist – for most of his career – about sets and numbers, while steadfastly refusing to accept the existence of propositions and properties or attributes. (*See* Quine, 1960, chs. 6, 7 and 1970, ch. 1).

(2) Realism and antirealism, though mutually exclusive, need not exhaust the possibilities. One may just be agnostic about whether or not there are abstracta (of some given kind). More interestingly, unwillingness to assert or deny the existence of abstracta might stem from a conviction that the issue is either hopelessly unclear or confused. CARNAP's view that philosophers' questions about the existence of numbers, propositions, etc., are not genuinely factual or "theoretical" questions at all, but misleadingly formulated "practical" questions – calling for a decision whether or not to adopt a certain "linguistic framework" rather than an answer assessable as true or false – can be seen as exemplifying the latter position. (*See* Carnap, 1950.)

(3) *Abstract* – as opposed to *concrete* – entities are commonly taken to be those, if any, which occupy neither space nor time. Thus they contrast both with physical entities, which occupy both SPACE AND TIME (e.g., tables, tennis matches, vapor trails, and more exotic entities like sub-atomic particles and force-fields), and with those entities, if any, which occupy time but not space (e.g., mental events, processes and states, on some dualist views), or space but not time (possible examples: the Greenwich Meridian, the North Pole, and spatial points and regions generally). This explanation is not unproblematic. While numbers and sets, and many other standard examples of the abstract, have neither spatial nor temporal location. It makes no sense to ask where the number 17 is, or when it came into existence, or how long it will last. But it is not clear that this holds for all abstract entities – one might, for example, argue that literary and musical works (as distinct from copies or performances) are abstract entities, but that they have not always existed – rather, they came into existence when first composed, and so are not wholly atemporal. We shall, however, assume its approximate correctness here. (*See* CONCRETE/ABSTRACT. For skepticism about the distinction, *see* LEWIS, 1986, pp. 81–6;

for discussion of difficulties in drawing it, *see* DUMMETT, 1973, ch. 14; Noonan, 1976; Hale, 1987, ch. 3).

(4) Realism is standardly taken to involve the further claim that the existence of abstracta is *objective*, this being understood in terms of *mind-independence* (*see* OBJECTIVITY). This is both natural and plausible, but not unproblematic. Indeed, the same examples illustrate the difficulty – novels and symphonies are (complex) abstract objects, but ones which would not have existed without a good deal of mental activity on the part of their makers. Perhaps the simplest way around this difficulty is to distinguish a strong form of realism which asserts that there are mind-independent abstract entities, and weaker forms which assert the existence of abstracta, but not their mind-independence. The strong realist thesis is itself open to more and less demanding interpretations, which differ over how mind-independence is understood. A minimum condition for the existence of certain entities to be mind-independent is that they would exist even if there were no minds, and so exist independently of our actual knowledge or beliefs about them. But we may distinguish an extreme form of realism, according to which the existence of abstracta is entirely independent, even in principle, of the possibility of our knowing of it, from more moderate forms which maintain that there are abstract entities which would exist even if there were no thinkers, but which accept an epistemological constraint to the effect that their existence must be detectable, at least in principle.

(5) Alexius MEINONG (1904) (*see* NON-EXISTENT OBJECTS) *denies* existence (German: *Existenz*) to abstract entities, but maintains that they have a different kind of being, sometimes called "subsistence" (German: *Bestand*). A closely related, but subtly different view – sometimes called "noneism" – is defended in Routley (1980) and Priest (2005). Meinong's doctrine is standardly classed as a kind of realism, but in our terms, Meinongian Realism counts, somewhat paradoxically, as a form of antirealism. It is tempting to suppose that Meinong is using the word "exists" in a restricted way, so as to apply only to what occupies space and time, or perhaps only to what is capable of causal interaction. If this were so, the disagreement between his Realism and realism as characterized here would be largely if not entirely verbal, at least as far as the ontological status of abstract entities is concerned. One might be similarly tempted to think that the disagreement between realists and antirealists in our sense is likewise a merely verbal one, in which antirealists are simply evincing a prejudice in favor of restricting application of the word "exists" to what is concrete. But while some antirealist polemics encourage such a view, the temptation should, in this case, be resisted. There is a genuine issue, and it concerns knowledge. If his position is to be taken seriously, the realist must claim that we can have at least some knowledge about some abstract entities. But then, given that such entities lack spatio-temporal location, and so must be incapable of standing in any causal or other natural relations to us, however remote, he faces a challenge to explain how knowledge about them is possible. We shall return to this issue.

2. *Some Realist Views*

One may, as noted, be a realist about one kind of abstract entities but not about others. We illustrate with three examples.

Universals One of the earliest and most famous realist doctrines is Plato's Theory of Forms, which asserts the existence such things as the Beautiful and the Just in themselves, over and above particular beautiful objects and just acts which, in Plato's view, more or less imperfectly exemplify them. Although Plato's usual term for the Forms (ειδοσ) is often translated as "Idea", it is clear that he takes them to be abstract entities existing independently both of our mental activity and of their instantiation in sensible particulars (*see* PLATO). In support of this view, it may be argued that there is something which different just acts, for example, have in common, in virtue of which they are all rightly said to be just, and that what they have in common does not depend for its existence upon any of those particular acts being performed. Each just act

occurs at a particular time in a particular place, but what they have in common has itself no spatio-temporal location. The detailed interpretation of Plato's theory and his arguments for it remain matters of scholarly controversy, but there is no doubt that his promulgation of the theory initiated a dispute over the nature and existence of universals – often conceived, in opposition to particulars, as entities such as general properties which may be wholly present at different times and places, or instantiated by many distinct particular objects – which has been actively pursued in much subsequent philosophy (*see* UNIVERSALS AND PARTICULARS).

Propositions Much as realists about universals argue for them by appealing to the existence of something common to different particular objects or events which all satisfy some general description or predicate (e.g., "blue", "square", "just", etc.), so some philosophers have argued that when different speakers or thinkers say or think, say, that 17^2+1 is even, or that Julius Caesar was assassinated, there is something common to their distinct linguistic performances or psychological acts or states. What they share is a common content – what is said or thought, as distinct from the saying or thinking of it. In other words, they all assert, or assent in thought to, the same proposition. Propositions in themselves, in contrast with the linguistic performances or psychological acts or states in which they expressed or encoded, have no spatial or temporal location, and hence are abstract objects (A classic statement of realism about propositions is BOLZANO, 1972).

Numbers, Sets and other Mathematical Entities The sentences of pure mathematics almost invariably involve expressions (simple or complex singular terms) whose ostensible role is to make reference to numbers of some kind, or sets, or other mathematical entities, along with quantifiers binding variables understood as ranging over such entities. Simple examples are:

$2 + 7 = 9$

Every set of real numbers which is bounded above by a real number has a least upper bound in the real numbers

For every set X there exists a set Y whose members are exactly the subsets of X

If these and similar sentences, taken at face-value, are true, then there must be numbers, sets, etc., to which they refer or over which they quantify. But such sentences *are* widely accepted as true, and are accepted as they stand, without benefit of some re-interpretation which dispels the appearance of reference to or quantification over numbers, sets, etc. Here we have the premises of an argument which makes at least a *prima facie* case for the existence of numbers, sets, and other mathematical entities – abstract entities, surely, if any are – and hence a case for realism (or Platonism, as it is often called – *see* PLATONISM) about mathematics.

3. *Antirealism*

Realism's traditional opponents have been nominalists (*see* NOMINALISM). Thus in the medieval dispute over universals, the nominalists insisted that there exist only particular entities, and that the application of the same general term (or name – hence the label "nominalism") to many distinct particulars does not require the existence of a common non-linguistic entity which is somehow present in each of them, but is sufficiently explained by reference to similarities between them. Likewise, in the modern dispute over the existence of abstract entities in mathematics, nominalists argue that the acceptance of mathematical theories involves no unavoidable commitment to the existence of numbers, functions, sets, or any other ostensibly abstract entities. Before we consider some of the main strategies by which nominalists have sought to avoid such commitment, we shall briefly review their reasons for thinking it is necessary or desirable to avoid it.

Nominalists have often recommended their rejection of abstracta on grounds of ontological economy, invoking the methodological maxim known as Ockham's Razor – *entia non sunt multiplicanda praeter necessitatem* – which may be glossed as asserting that we should not postulate kinds of entity beyond what is necessary (*see* OCKHAM). Although a popular ploy, this is problematic

for at least two reasons. First, it gives a clear directive only when accompanied by some answer to the obvious question: "Necessary for what?" The equally obvious answer is: "Necessary to account for all the (agreed) facts", but it is doubtful that there is sufficient agreement here to enable the nominalist to cut away abstracta as unnecessary. The realist is likely to suppose that the relevant facts include facts of mathematics which, taken at face value, *do* require the existence of numbers, sets, etc. But second, even if the facts in need of explanation can be restricted, without begging the question, to facts about the concrete, it is still unclear that the nominalist will be in position to wield the razor to advantage, since it may be argued that those facts admit of no satisfactory explanation without the aid of scientific (and especially physical) theories which make indispensable use of mathematics. This – often called the Quine–Putnam indispensability argument – receives its clearest formulation in PUTNAM (1971). Since theories (especially mathematical theories) ostensibly involving reference to abstracta appear to play an indispensable rôle in our intellectual economy, nominalists can scarcely afford simply to reject them outright; rather, they must explain how we may justifiably retain such theories, without offending against nominalistic scruples.

The standard nominalist response has been to seek ways of paraphrasing or re-interpreting problematic statements and theories in nominalistically acceptable terms – with the aim of showing that their apparent reference to and quantification over abstract entities is unnecessary or merely apparent. This strategy has met with limited success. The difficulties can be well illustrated by reference to arithmetic. Consider first simple equations, such as "2 + 3 = 5". As a step towards eliminating its apparent reference to numbers, we may paraphrase it along the lines: "If there are exactly two Fs and exactly three Gs and no Fs are Gs, then there are exactly five F-or-Gs" (in symbols: $(\exists_2 xFx \wedge \exists_3 yGy \wedge \neg\exists x(Fx \wedge Gx)) \rightarrow \exists_5 x(Fx \vee Gx))$. Although this still contains number words, they occur only in the context of numerically definite quantifications like "there are exactly two Fs" ($\exists_2 xFx$). These are logically equivalent to sentences involving no number words at all, such as "there is something which is F and something else which is F and any F is identical with one or other of these things" ($\exists x\exists y(x{\neq}y \wedge \forall z(Fz{\leftrightarrow} z{=}x \vee z{=}y))$). Thus at some cost in length and readability, we may be able to reduce "2=3=5" to something nominalistically acceptable. But even if this kind of paraphrase works for simple equations, it plainly won't work for general arithmetical statements, such as $\forall a\forall b\ (a + b = b + a)$, in which we quantify over numbers, without mentioning any in particular. Thus unless virtually the whole of arithmetic is to lie beyond the nominalist's reach, additional and more widely applicable methods of paraphrase or re-interpretation will be needed.

Eliminative structuralism offers a more promising strategy. On this account, arithmetic is *not* a theory about a particular infinite sequence of abstract objects – the numbers 0,1,2,3, . . . – but gives completely *general* information about those objects, *if any*, which exemplify a certain structure (viz. being a sequence having a first term, and for each term, a unique next term, and so no end of terms – progressions, or ω-sequences, in the usual jargon). Since, on this re-interpretation, no arithmetic sentences assert the existence of any objects, they are all nominalistically acceptable. A well-known difficulty is that unless there exists at least one ω-sequence, the eliminative structuralist's translations of all arithmetic sentences, including those of false ones like 2+3=6, come out true. This leaves the nominalist facing a dilemma: to avoid this disaster, she must assert the existence of an ω-sequence – but if she asserts that there infinitely many abstract objects, she abandons nominalism, while if she asserts that there are infinitely many concrete objects, the viability of her translation-scheme depends upon an empirical hypothesis, and one which may very well be false. Perhaps, as Hellman (1989) argues, this dilemma can be avoided by strengthening the structuralist translations

so that they make claims about what *necessarily* holds of any ω-sequence – for then the nominalist need only assert the *possible* existence of an ω-sequence to avoid disaster, and perhaps the claim that there could be an ω-sequence is nominalistically unproblematic and otherwise acceptable. However, even if a nominalist version of arithmetic can be salvaged in this way, it is doubtful whether the strategy can be extended to more powerful mathematical theories such as set theory, since the needed possible existence claim would amount to the claim that there could be a *concrete* model of transfinite set theory, and this is surely false.

Following a more radical course, Hartry FIELD (*see* Field, 1980, 1989) has argued that nominalists can *deny* that mathematical theories are *true*, thereby avoiding commitment to their abstract ontology, but still *accept* them provided they are *conservative* in the sense that their conjunction with non-mathematical (e.g., physical) theories entails no claims about non-mathematical entities which are not logical consequences of those non-mathematical theories by themselves. Conservativeness in this sense, like logical consistency, does not require truth – a theory can be conservative without being true. The important uses of mathematics in science, Field holds, are two: we use it to deduce the consequences of scientific theories, and we use it, especially in physics, in actually formulating such theories. The assumption that standard mathematics is conservative, Field argues, is enough to justify its use in deducing, and with the help of this assumption, we can, he thinks, show that there are acceptable nominalistic reformulations of such theories. Field's view has attracted a barrage of objections, both technical and philosophical. Several critics have questioned whether Field's reformulations of scientific theories really are nominalistically acceptable. Others have argued that he is committed to the implausible view that while there exist no numbers or sets, their non-existence is a merely contingent matter. (*See* Maddy, 1980; Chihara, 1990; Hale and Wright, 1992; Burgess and Rosen, 1997).

4. *Vehicles of ontological commitment – reference and quantification*

It was claimed above that a *sufficient* condition for the existence of objects of a given kind, F, is the occurrence in true statements of expressions functioning as singular terms which, if they refer at all, refer to Fs. Such terms are, we might say, vehicles of ontological commitment. It might be objected that the suggested condition cannot be sufficient as it stands, and that we should additionally require that the relevant singular terms be ineliminable by reductive paraphrase of the sort orthodox nominalists have sought to supply. But this objection is confused. Accepting as true statements in which certain expressions function as singular terms commits us to the existence of corresponding objects, simply because those statements cannot be true unless their ingredient expressions discharge their semantic functions, and the semantic function of singular terms is to pick out objects.

Antirealists may agree, but object that this misses the real point, which is that if statements apparently involving singular terms for abstract objects can indeed be replaced by equivalent statements which do not, this shows that those terms are not genuine singular terms at all, and that the original statements, contrary to first appearances, involve no commitment to such objects. This antirealist counter assumes that if statements apparently involving ontological commitment to Fs are equivalent to other statements apparently free of any such commitment, it is the latter statements which should be reckoned as truly reflecting our ontological commitments, not the former. But why? An equivalence, as ALSTON (1958) points out in a perceptive discussion of the issue, is just that – what is shows, by itself, is only that if either of the two kinds of statement involves a commitment to Fs, then both do. But to get to the conclusion that statements of the first sort involve no genuine reference (and hence commitment) to Fs, we need a further premise – one providing a reason to regard the appearances presented by statements of that sort as misleading, in

contrast with those presented by statements of the other sort. Suppose we could introduce terms for the directions of straight lines by means of the *Direction Equivalence*:

The direction of line a = the direction of line b iff lines a and b are parallel – the idea being to establish a use for such terms by fixing the truth-conditions of identity statements involving them. (*See* FREGE, 1884, §64.) The nominalist will regard the equivalence as revealing that any apparent commitment to the existence of abstract objects carried by talk of directions is merely apparent. The realist will instead regard it as disclosing an unobvious commitment to the existence of directions implicit in talk of parallelism among lines. Of course, the realist must agree that one could possess the concepts of straight line and parallelism without having that of direction – indeed, one must be able to do so, if the latter concept is to be explained by means of the Direction Equivalence. His claim is that the commitment to directions is implicit in the sense that, once one has acquired the concept of direction in this way, one cannot consistently hold that there are straight lines but no directions. (For further discussion, *see* WRIGHT, 1983, §§5,10.) The realist claims we should take apparent reference to abstracta at face value, in the absence of compelling reason to do otherwise. Resolution of the issue in favor of an ontologically reductive interpretation of such equivalences – and so in favor of the antirealist – requires making a case that there is compelling reason to do otherwise. We shall return to this question.

Our proposed sufficient condition for the existence of Fs is clearly not a *necessary* condition. It may be that there are Fs whose existence we suspect not, and of which, therefore, we do not speak. Perhaps, indeed, we have no concept of them. Nor, evidently, is a readiness to make statements involving singular terms for Fs needed for a commitment to their existence. For without employing any words which purport reference to particular Fs, we may simply assert that there are Fs, or more generally, assert some quantified statement whose truth requires their existence. Roughly, quantification over Fs is an alternative vehicle of ontological commitment to Fs. Quine, famously, took it to be the *sole* vehicle:

> The objects whose existence is implied in our discourse are finally just the objects which must, for the truth of our assertions, be . . . reckoned into the totality of objects over which our variables of quantification range. To be is to be the value of a variable. (Quine, 1952, §37)

Quine sees quantification as *the* vehicle of ontological commitment because he assumes that only *ineliminable* occurrences of singular terms would *distinctively* carry ontological commitment, and believes that there are no such terms, i.e., that singular terms are everywhere eliminable. We may, he argues, always eliminate them by paraphrase using just general terms (predicates) and quantification, either by the technique of RUSSELL's Theory of Definite Descriptions (coupled with his doctrine that ordinary proper names are "abbreviated" descriptions) or, if necessary, by an extension of it due to Quine himself whereby we may replace any proper name by a corresponding predicate understood as applying to that object, if any, the name names – thus "Socrates drinks", for example, may be paraphrased as "$\exists$x(x socratizes & x drinks)". Quine is also taking it for granted that predicates or general terms carry no commitment to corresponding entities. If this assumption – to which we shall need to return – were granted, it would be at least plausible that quantification over Fs is the essential mark of commitment to their existence.

However, while Quine's eliminability thesis is, in one way, beyond dispute, its significance is not. We may agree that, starting from a base language containing singular terms, we could employ Quine's recipe to construct a language in which all such terms were replaced by corresponding predicates, but deny this purely syntactical manoeuvre has any semantic or, more widely, philosophical significance. It is quite unclear how one might learn the use or satisfaction conditions of Quine's replacement predicates, in the absence of any means of making singular reference to the objects

which, if any, uniquely satisfy them. Relatedly, it does not seem one could explain the truth-conditions of quantified sentences of Quine's replacement language without treating variables as, in effect, functioning as temporary names of objects in the domain of quantification. (*See* Dummett, 1973, pp. 223–6, 476–80.)

The Access Problem

Realists need to explain how we can know about the abstract entities whose existence they assert – how we can know that there are such things at all, and how we can know truths about them. The problem of providing such an explanation is part of what I shall call the *access problem*. It is the fundamental problem for realism. If realists could solve it, it is difficult to see what, other than prejudice, would stand in the way of acceptance of their view. If, on the other hand, it could be shown that they *cannot* solve it, that would be a decisive objection, and would encourage, or even enforce, an ontologically reductive reading of the kind of equivalences between statements ostensibly about abstracta and others apparently free of commitment to their existence discussed in the preceding section. In the absence of at least the outlines of a solution, or reason to believe one can be found, it is hard to take realism seriously – ontology without epistemology is just idle speculation. (*See* Hart, 1979; Bell, 1979).

Why do we – or might we – find the idea that we may have knowledge about abstract objects so baffling? In explaining how we know much of what we know, we appeal to causal connections, such as those involved in perception. This may encourage acceptance of a broadly causal theory of knowledge – one which sees basic bits of knowledge as involving a suitable causal connection between knowers and the known truths, and other knowledge as arising from this basis by a more or less complicated process of inference. Then, given that abstract objects stand in no spatial or temporal relations with us, and so in no causal relations, it may seem not just that knowledge about them eludes explanation, but that there can be no such knowledge. (*See* Benacerraf, 1973; Steiner, 1973; Kitcher, 1978). As against this, it may be claimed that even a broadly causal theory is open to objection on independent grounds; in particular, such a theory would seem directly to rule a priori knowledge – and while there is certainly a serious problem in explaining how such knowledge is possible, it does not seem that its impossibility should be so easily established. However, as Field (1989, pp. 25–7, 230–9) and others (Hart, 1977; Maddy, 1990, pp. 42–5) have pointed out, doubts about the capacity of realism to deliver a credible epistemology do not have to be grounded in the adoption of a specifically causal *analysis* of knowledge. For even if a causal constraint is not written into the analysis, the problem of explaining how we can acquire knowledge, or reliably form true beliefs about abstract objects, remains.

Epistemological perplexity about, and consequent suspicion of, abstract entities has other and more general sources, besides causalist or, more generally, reliabilist thinking in epistemology, which arguably obstruct progress on the access problem.

One is that we tend to operate with a wholly *negative* conception of abstract objects as "outside" space and time. This characterization is obviously metaphorical, as well as negative – there is, literally, nowhere outside space and time. But this in itself need not be particularly damaging, so long as we remind ourselves, when necessary – that is, when we feel tempted to think of abstract objects as "in" some queer sort of limbo – of the literal content of the metaphor: roughly, that it makes no sense to ask where an abstract object is, or when it came into existence, or how long it will last. It is, rather, the negative aspect of the characterization that impedes constructive thought. Of course, it is true that abstract objects aren't located in space or time. And it may be said that since that it enough to ensure that there is an apparently intractable problem about how spatio-temporally located knowers could know of their existence or know anything about them, it is pointless exercising ourselves over what more positive

characterization, if any, they can be given. But that just misses the present point: if we focus exclusively on what abstract objects are *not*, with no thought about what they *are* or might be supposed to *be*, we can scarcely expect anything but intellectual paralysis when we try to consider how we might get to know about them.

The second factor is the idea that knowledge of truths about objects of any kind must involve "contact" with those objects. If "contact" is taken literally, so as to require some sort of physical connection or interaction – perhaps of the sort that occurs in normal sense perception, or even something more indirect – the idea is obviously inimical to realism, but equally not obviously one that must be accepted. Of course, if it is given a sufficiently attenuated (and perhaps unavoidably metaphorical) construal, so that possession of any sort of identifying knowledge of an object suffices for contact, the idea reduces, near enough, to a truism – one can hardly be credited with knowledge of truths about objects unless one knows which objects are in question – and it need then cause the platonist no trouble, unless it is coupled with the further idea that such "contact" is presupposed by and must be already in place before any knowledge of truths about objects can be had (cf. Russell's famous principle that "Every proposition which we can understand must be composed wholly of constituents with which we are acquainted" (*see* Russell, 1912, chs. 4, 5).

Once we become locked into thinking about the access problem within this straitjacket, we can hardly avoid the further thought – that the access problem is not just a problem about how we can *know* anything about abstract objects, but goes wider and deeper: how can we even so much as *think about them* at all.

Critics of realism may see this as just so much more grist to their mill: realism is in trouble on *two* counts, not just *one*, because it obstructs *both* a satisfactory epistemology *and* a workable theory of reference (cf. Benacerraf, 1973, p. 412; Field, 1989, p. 68). But this way of putting the difficulty obscures an important connection. The right way to put the objection is like this: even if one *could* give a realist account of the truth conditions for mathematical statements (or any other class of statements supposed about abstract objects), it would be impossible to explain how such statements, so understood, could be known or reasonably believed; but in fact one cannot even give such a semantical account, since one cannot even so much as make reference to "objects" of the sort such an account takes them to be about – and if one cannot do that much, one cannot so much as state realist truth-conditions. It helps to recast the objection in this way, because doing so gives a clearer view of the structure of the task that must be addressed by a defensible form of realism. The *fundamental* part of the access problem is not the *knowledge problem* (i.e., how, given that certain statements (e.g., mathematical ones) are about abstract objects, we could know them to be true), but the *reference problem* (i.e., how they could be about such objects in the first place).

That, then, is the problem the realist should tackle first. Although solving the reference problem is merely a necessary, and not a sufficient, condition for a solution to the knowledge problem, one might expect a good solution to the former to suggest how best to approach the latter. But how, if at all, may realists solve the reference problem? In my view (for a concise statement, *see* Hale and Wright, 2002, sect. 5), their best hope lies in rejecting the assumption that an ability to engage in identifying reference to, or thought about, abstract objects is a *precondition* for understanding statements about them, as is suggested by the "contact" model and Russell's Acquaintance Principle (*see* ACQUAINTANCE). Positively, they should argue that concepts of kinds of abstract object may be introduced by fixing the truth-conditions of complete sentences involving terms for them, in accordance with Frege's Context Principle ("Only in the context of proposition does a word mean anything" – cf. Frege, 1884, §62). More specifically, they may then deploy what have come to be known as "abstraction principles" as a means of explaining both how terms for abstract objects are to be understood

and how basic truths about them may be known a priori. Examples are the Direction Equivalence (*see* above, sect. 4) and *Hume's Principle*:

> The number of Fs = the number of Gs iff the Fs correspond one–one with the Gs

Whether the access problem can be solved in this, or some other way, is a matter of currently active debate.

BIBLIOGRAPHY

Alston, William: "Ontological Commitments," *Philosophical Studies* 9 (1958), 8–17.

Bell, David: "The Epistemology of Abstract Objects," *Proceedings of the Aristotelian Society*, suppl. vol. 53 (1979), 135–52.

Benacerraf, Paul: "Mathematical Truth," *Journal of Philosophy* 70 (1973), 661–80.

Bolzano, Bernard: *Theory of Science*, ed. and trans. Rolf George (Oxford: Blackwell, 1972). From *Wissenschaftslehre* (Sulzbach, 1837).

Burgess, John P. and Rosen, Gideon: *A Subject With No Object: Strategies for Nominalistic Interpretation of Mathematics* (Oxford: Clarendon Press, 1997).

Carnap, Rudolf: "Empiricism, Semantics, and Ontology," *Revue Internationale de Philosophie* 4 (1950), 20–40.

Chihara, Charles: *Constructibility and Mathematical Existence* (Oxford: Clarendon Press, 1990).

Dummett, Michael: *Frege: Philosophy of Language* (London: Duckworth, 1973).

Field, Hartry: *Realism, Mathematics, and Modality* (Oxford: Blackwell, 1989).

Field, Hartry: *Science Without Numbers* (Oxford: Blackwell, 1980).

Frege, Gottlob: *Die Grundlagen der Arithmetik* (Breslau, Poland: Wilhelm Koebner, 1884); trans. into English by J.L. Austin as *The Foundations of Arithmetic* (Oxford: Blackwell, 1959).

Hale, Bob: *Abstract Objects* (Oxford: Blackwell, 1987).

Hale, Bob and Wright, Crispin: "Benacerraf's Dilemma Revisited," *European Journal of Philosophy* 10 (2002), 101–29.

Hale, Bob and Wright, Crispin: "Nominalism and the Contingency of Abstract Objects," *Journal of Philosophy* 89:3 (1992), 111–35.

Hart, W.D.: "The Epistemology of Abstract Objects," *Proceedings of the Aristotelian Society*, suppl. vol. 53 (1979), 153–65.

Hart, W.D.: Review of Mark Steiner, *Mathematical Knowledge* (Ithaca, NY: Cornell University Press, 1975), in *Journal of Philosophy* 74 (1997), 118–29.

Hellman, Geoffrey: *Mathematics Without Numbers* (Oxford: Clarendon Press, 1989).

Kitcher, Philip: "The Plight of the Platonist," *Noûs* 12 (1978), 119–36.

Lewis, David: *On the Plurality of Worlds* (Oxford: Blackwell, 1986).

Maddy, Penelope: *Naturalism in Mathematics* (Oxford: Clarendon Press, 1990).

Maddy, Penelope: *Realism in Mathematics* (Oxford: Clarendon Press, 1980).

Meinong, Alexius: "The Theory of Objects," in R. Chisholm, R., ed. *Realism and the Background to Phenomenology* (London: Allen & Unwin 1960; originally published 1904).

Noonan, Harold: "Dummett on Abstract Objects," *Analysis* 36:2 (1976), 49–54.

Priest, Graham: *Towards Non-Being: The Logic and Metaphysics of Intentionality* (Oxford: Clarendon Press, 2005).

Putnam, Hilary: "Philosophy of Logic," New York: Harper & Row, 1971; repr. in Putnam's *Philosophical Papers Volume 1* (Cambridge: Cambridge University Press 1975).

Quine, W.V.: *Methods of Logic* (London: Routledge and Kegan Paul, 1952). Quine, W.V.: *Philosophy of Logic* (Englewood Cliffs, NJ: Prentice-Hall, 1970).

Quine, W.V.: *Word & Object* (Cambridge, MA: MIT Press, 1960).

Routley, Richard: *Exploring Meinong's Jungle and Beyond* (Canberra: Australian National University, 1980).

Russell, Bertrand: *The Problems of Philosophy* (London: Oxford University Press, 1912).

Steiner, Mark: "Platonism and the Causal Theory of Knowledge," *Journal of Philosophy* 70 (1973), 57–66.

Wright, Crispin: *Frege's Conception of Numbers as Objects* (Aberdeen: Aberdeen University Press, 1983).

BOB HALE

Space and Time

This article discusses the following issues about space and time: whether they are absolute or relative, whether they depend on minds, what their topological and metrical structures may be, MCTAGGART's argument against the reality of time, the ensuing split between static and dynamic theories of time, problems with presentism, and the possibility of time travel. Our opening questions are posed in the following query from KANT:

> What, then, are space and time? Are they real existences? Are they only determinations or relations of things, yet such as would belong to things even if they were not intuited? Or are space and time such that they belong only to the form of intuition, and therefore to the subjective constitution of our mind, apart from which they could not be ascribed to anything whatsoever? (A23/B37)

ABSOLUTE OR RELATIVE?

NEWTON regarded space as a real existence – a vast aetherial container without walls, in which everything else that exists lives and moves and has its being. LEIBNIZ believed to the contrary that space is not a genuine entity, but a mere *façon de parler*; he held that all talk of space is replaceable by talk of material things and their relations to one another. For example, to say that a thing has "changed its place" is merely to say that it has changed its distance or direction from some other thing chosen as a reference object. This is the issue that divides partisans of absolute or substantival theories of space on the one hand from defenders of relative or relational theories on the other.

To test his or her allegiance on this issue, the reader should answer the following question: if the only material thing in existence were a single particle, would it make sense to say that it is moving? Leibniz would say no, since motion for him consists in change of relations (e.g., of distance) among two or more material things. Newton would say yes, since even in the absence of other material things, the particle could be moving from one cell to another of space itself.

Newton argued for the existence of substantival space with a famous thought experiment. Imagine a bucket suspended from a rope and filled with water. The rope is twisted and allowed to unwind, causing the bucket to spin. At first the bucket moves relative to the water, the water not yet having begun to partake of the bucket's motion, but eventually friction causes the water to rotate as well, and indeed to "catch up" with the bucket so that there is no longer any relative motion between water and bucket. By the time this happens, something else happens as well: the surface of the water has become concave, the water edging up the sides of the bucket. This is explained in Newtonian mechanics as a centrifugal-force effect, similar to what happens when amusement park riders are pinned to the side of a rotating bottomless drum. Newton's argument now proceeds as follows:

1. There is a time at which the water displays centrifugal-force effects, but is not moving relative to the bucket – or any other material thing. (Why not relative to the ceiling, you ask? That is why the experiment is a thought experiment: we are to imagine it performed in a universe with no objects besides bucket, water, and rope.)
2. All centrifugal-force effects are induced by rotational motion
3. Therefore, there is a time at which the water is moving, but not relative to any material thing (from 1 and 2).
4. Motion that is not relative to any material thing is absolute motion, that is, motion with respect to space itself.
5. Therefore, the water is moving with respect to space itself (from 3 and 4) – which must therefore exist.

Newton thus argues that accelerated motion (the water's constant change of direction) reveals itself in its effects and proves the existence of space, the existence of which then grounds absolute *uniform* (non-accelerated) motions, even though the latter do not manifest themselves.

BERKELEY and Leibniz objected to the conclusion of Newton's argument, but without making clear which premise they thought wrong. Mach objected to premise 1, claiming that we simply do not know how the water would behave in a universe devoid of ceiling and stars (as though no physicist ever extrapolated his laws to hypothetical situations, such as frictionless planes!). A generally overlooked response challenges premise 4: perhaps motion is *really* absolute, that is, not a change in relation to anything else at all, be it matter or space. The possibility of this last response shows that we should separate two issues that can be posed using the "absolute vs. relative" formula: is space a substance or a system of relations, and are motion, size, and various other spatial commodities absolute (intrinsic) or relational?

Leibniz argued that space is a pseudo-entity because its existence would generate distinctions without a difference or, more precisely, exceptions to his principle of the IDENTITY OF INDISCERNIBLES. Let w and w′ be two universes just alike in how all material things are related to one another, but differing in the alleged respect that in w′ the entire material cosmos has been moved six miles to the east or rotated through some angle. Leibniz's argument then proceeds as follows:

1. If there were such a thing as substantival space, w would be distinct from w′.
2. But w and w′ are indiscernible – they share all their properties.
3. Things that are indiscernible are identical. Putting it the other way around, any two distinct things must differ in at least one property.
4. Hence, w = w′ after all (from 2 and 3).
5. Therefore, there is no such thing as substantival space (from 1 and 4).

To evaluate this argument, we need to distinguish two kinds of properties. A property is *pure* if its being exemplified does not depend on the existence of any specific individual and *impure* otherwise. Examples of pure properties are being red (which is pure and intrinsic) and being next to something red (pure and relational); examples of impure properties are being Fred (impure and intrinsic) and being married to Fred (impure and relational). When Leibniz affirms premise 2, he must mean that w and w′ differ in no pure property, for Newtonians could certainly maintain that w and w′ are distinguished by the fact that w is such that part of the cosmos occupies cell 233 (an impure property), whereas in w′, cell 233 is empty. But that means when we get to premise 3, Leibniz must advance his Identity of Indiscernibles principle in the following form: any two things must differ in at least one *pure* property (and not merely in such properties as being identical with *this thing*). Leibniz no doubt did wish to affirm the principle in the required form, but if so, it is open to counterexamples. Is it not conceivable that there be two spheres the same in color, shape, composition, and every other pure property you care to think of?

The substantival vs. relational issue carries over to time. For Newton, time "flows equably without regard to anything external;" for Leibniz, time is nothing over and above the sequence of events said to be in time. Newton (but not Leibniz) can make sense of the idea that the entire history of the world (comprising the same events as now) might have begun earlier than it did.

REAL OR IDEAL?

Another issue about space and time is whether they are ideal, that is, dependent for their existence on minds. The most famous idealist about space and time in western thought is Kant. Kant began his intellectual career as a Leibnizian, but was briefly converted to Newton's view by considerations about "incongruent counterparts" – objects that come in mirror image forms, like left and right human hands. Kant thought the difference between incongruent counterparts could not be explicated using only relationist resources, but had to consist in the differing relations of the objects to space itself. By the time he wrote the *Critique of Pure Reason*, however, Kant had come around to the third of the positions in the quotation above: space and time are merely "forms of intuition," that is, ways in which human

beings order and arrange the things they perceive; they are not features of things in themselves, or things as they exist outside the mind.

A characteristically Kantian reason for believing that space is ideal is that no other hypothesis accounts for our knowledge of geometry. Kant thought that geometry was a body of synthetic and a priori truth – a priori in that it is known in advance of experience, yet synthetic in that it is not validated just by logic or the meanings of our concepts. How can that be? How can we know even before we encounter them that cubes on Mars will have 12 edges? Kant's answer is that (i) our form of intuition makes us incapable of intuiting (perceiving or imagining) any cubes that do not have 12 edges and (ii) as prescribed by idealism, no cubes or spatial objects exist anywhere except those that satisfy the conditions of our intuiting them. Thus all cubes everywhere have 12 edges and the other properties imposed on them by our Euclidean form of intuition.

Kant thought the ideality of space and time was further confirmed by the antinomies – pairs of opposed propositions in which one or the other must be true if space and time exist outside the mind, but both of which are impossible. For example, does the world have a beginning in time, or is it infinite in its past duration? If things in time were things in themselves, one of these alternatives would have to be true, yet both of them boggle the mind. *No beginning* would mean an infinity of events already elapsed, which Kant thought impossible because it would involve a "completed infinity." (Think of WITTGENSTEIN's example of the man we find saying, ". . . −5, −4, −3, −2, −1; whew! I just finished counting through all the negative integers.") *A beginning* would mean an event for which there could not possibly be a sufficient reason – a blow to rationalist aspirations, if not the outright impossibility Kant seemed to think it was. Kant's solution was to hold that past events exist only in present or future memories or other evidence of them (for example, yet-to-be-perceived cosmic radiation). He thought this opened the possibility that the world's history is *potentially* infinite – always extendable further into the past through our future discoveries – but neither actually finite nor actually infinite.

STRUCTURAL QUESTIONS

The next group of questions about space and time (or spacetime, in the Minkowskian melding of them) concerns their (or its) metrical and topological structure. Are space and time infinitely divisible, or are there smallest units? (Zeno's paradoxes of motion are sometimes seen as set up so that the first two apply if space and time are infinitely divisible and the second two if space and time are quantized.) Does space obey the laws of Euclidean geometry or those of one of the non-Euclidean geometries known to be consistent since the nineteenth century? How many dimensions does space have? Could time have a beginning or an end? Must time be unilinear, or might it branch into multiple paths or close back upon itself in a loop?

The dimensionality of space is representative of such questions. We all know about three dimensions of space – a line possesses one dimension, a plane two, and a solid three. What would it mean for space to have a fourth dimension? (We are talking now of a fourth *spatial* dimension, not time, even though time is sometimes considered as a fourth dimension.) Galileo offered one criterion: to say that space has n dimensions is to say that n mutually perpendicular lines (but no more) can meet in a single point. If our space were four-dimensional, a line could enter the corner of my desktop at right angles to each of its three edges. Poincaré offered another criterion: points are zero-dimensional, and an entity is n-dimensional iff n is the lowest number such that any two points of the entity may be separated from each other by an entity of $n - 1$ dimensions. Thus, a line has one dimension, because any two points of it can be separated from each other by an intervening entity of zero dimensions (another point); a plane has two dimensions, because any two points within it may be separated by a circle enclosing one of them or a line

running all the way across the plane between them; and so on. It is a consequence of this criterion that in a four-dimensional space, a two-dimensional entity would not suffice to separate one point from another. Thus a spherical shell enclosing point A but not point B would not suffice to separate A from B – you could get from A to B without penetrating the shell.

Such things defy visualization in a way that makes some people want to declare them impossible. Those so inclined should read E.A. Abbott's Victorian classic *Flatland*, in which the author describes a world of two-dimensional beings who are incapable of rising out of their plane or visualizing anything beyond it. A Flatlander may be imprisoned simply by enclosing him within a circle or a polygon. Could a Flatlander but jump over the walls of his prison, he would be free, but he is incapable even of conceiving such a motion – as we are of any path from the interior to the exterior of a spherical shell that does not pass through the shell. The exhortation "Upwards, not northwards!" falls on the Flatlander's ears as nonsense. Abbott's intent, of course, is to soften us up for the possibility that our own resistance to a fourth dimension may be as provincial as that of the Flatlanders to a third.

Questions about the structure of space and time give rise to meta-questions about proper jurisdiction – who is to answer them, and how? A traditional view is that space and time necessarily possess whatever structure they do, and that it ought to be ascertainable a priori what this structure is. Kant, for example, certainly believed that space is necessarily three-dimensional and Euclidean. The prevalent contemporary view is that space and time have their structures contingently, and that it is only through the best science of the day that we can reach any reasonable opinion concerning what these structures are. This view was given impetus by Einstein's use of a non-Euclidean geometry in conjunction with the General Theory of Relativity to explain gravitation; it is further exemplified in the work of those physicists in search of a "theory of everything" who posit a space of 11 dimensions.

A view that lies between the traditional and the contemporary views is the conventionalism of Poincaré. Poincaré thought that all the empirical data accommodated by non-Euclidean geometry plus standard physical theory could equally well be accommodated by Euclidean geometry together with non-standard physical theory. For example, measurements apparently indicating that the ratio of circles to their diameters does not have the familiar value of π could be accommodated by a non-Euclidean geometry in which this ratio is indeed other than π, but they could also be accommodated by positing a heat gradient that causes our yardsticks to expand when laid along the diameter though not when laid along the circumference. We could thus always choose to describe our world in Euclidean terms by complicating our physics. This position is at odds with a hardy empiricism, in so far as it denies that empirical results can settle the structure of space, but it is also at odds with an ambitious *a priorism*, in so far as it denies that decisions in favor of Euclid are determinations of independent fact.

QUESTIONS ABOUT TIME

For issues specifically about time, the best point of departure is McTaggart's famous argument of 1908 that time is unreal. Though few have accepted the conclusion of this argument, nearly all students of time have taken over the distinctions McTaggart employed in formulating it.

McTaggart's fundamental distinction is between the A-series and the B-series. An A-series is a series of events or moments possessing the characteristics of being past (in varying degrees), present, or future; call these the A-characteristics. The B-series is a series of events or moments standing in the relations of earlier-than, later-than, and simultaneous with; call these the B-relations. The chief difference McTaggart notes between the A-characteristics and the B-relations is that the former are transient while the latter are permanent: "If M is ever earlier than N, it is always earlier. But an event, which is now present, was future, and will be past" (LePoidevin and MacBeath,

1993, p. 24). In the ordinary way of thinking about time, McTaggart believes, an event becomes increasingly less future, is momentarily present, and then slides ever farther into the past. Yet all the while its B-relations to other events (e.g., its following the Battle of Waterloo and preceding the first landing on the moon) are fixed.

McTaggart's overall argument against the reality of time may be stated quite briefly: (I) time essentially involves an A-series; (II) any A series involves a contradiction; therefore, (III) therefore, time is unreal. Behind each main premise is a subsidiary argument. The argument behind premise I is this:

1. There can be no time without change.
2. There can be no change without an A-series.
3. Therefore, there can be no time without an A-series.

Both premises in this argument have been the subject of interesting debate, but our focus here will be on the argument behind main premise II, which runs thus:

1. The A-characteristics are mutually incompatible, yet
2. Every event in any A-series must have all of them, so
3. Any A-series involves a contradiction.

McTaggart immediately anticipates an objection the reader will have to premise 2: it is not true that any event must have all the A characteristics at once, but only that it must have them successively. An event that is now present is not *also* past and future; rather, it *was* future and *will* be past. In reply, McTaggart claims that this attempt to avoid the contradiction he alleges only raises it anew. What, he asks, is meant by tensed verb forms such as "was" and "will be"? His answer may be given in the schema

> S {was, is now, will be} P iff for some moment m, S *has* P at m & m *is* {past, present, future}

where the italicized verbs are meant to be tenseless. He thus believes that tense can be reduced to A-characteristics and tenseless copulas. If this is right, then in saying that an event has been future and will be past, we are introducing a new A-series, this time of moments. And this brings back our contradiction, because every moment, like every event, is past, present, and future. If we try to get rid of the contradiction by saying of moments what we said earlier about events, our statement "means that the moment in question is future at a present moment, and will be present and past at different moments of future time. This, of course, is the same difficulty over again. And so on infinitely" (LePoidevin and MacBeath, 1993, p. 33).

Why is McTaggart so convinced that there is a contradiction in the A-series and a regress in any attempt to remove it? His thought on these matters can be made more understandable by presenting it with the help of a metaphor. He begins by supposing that the whole of history is laid out in a block comprising the B-series. He notes that in such a series, there is no change and therefore no time, all events simply sitting there alongside one another on the B-axis. What can add time to such a universe? We must bring in the A-characteristics, letting the spotlight of presentness wash along the series in the direction from earlier to later. But wait! If the spotlight illuminates event *e* before it illuminates event *f*, then the events of *e's being present* and *f's being present* are both there on the B-axis, permanently related by the relation of earlier-than. Similarly, if the shadow of pastness falls on *e* before it falls on *f*, then *e's being past* and *f's being past* permanently stand in the B-relation of earlier-than and are thus always there on the B-axis. What we are saying implies that that *e* and *f* are both always past and always present – surely a contradiction, just as McTaggart alleges. If we seek to remove the contradiction by saying that the spotlight of the present falls on *e's being present* before it falls on *f's being present*, we are only embarking on a useless regress – again just as McTaggart alleges.

As noted above, few besides McTaggart have accepted his argument *in toto*, but many have accepted one half or the other.

This gives rise to a great divide in the philosophy of time. One side accepts his first main premise while rejecting the second: the A-characteristics (or some surrogate for them) are indeed essential to time, but there is nothing wrong with that. The other side accepts his second main premise while rejecting the first: there is indeed a defect in the A-series, but a B-series by itself is all you need to have time. For obvious reasons, these two responses to McTaggart are often called "the A theory" and "the B theory," though the names can be misleading.

There is an entire cluster of doctrines that tend to go together under the banner of the A theory and another opposing cluster under the banner of the B theory. (Other labels for the two sides are the dynamic versus the static theory and the theory of passage or becoming versus the theory of the four-dimensional manifold.) The rival doctrines may be tabulated as follows:

The A Theory (Dynamic Time)	The B Theory (Static Time)
A1. Tense is an irreducible and indispensable feature of thought and language, reflecting a genuine feature of reality. Corollary: some propositions change in truth value with the passage of time.	B1. Tense is reducible or eliminable; reality is adequately describable without it. Corollary: every true proposition is timelessly true.
A2. The A-characteristics are successively possessed by all events, and they are not reducible to the B-relations.	B2. The A-characteristics are either delusive or reducible to the B-relations.
A3. The present is ontologically privileged: things present have a reality not belonging to things past or future.	B3. Past, present, and future are ontologically on a par: things past and future are no less real than things present.
A4. The future is open or indeterminate: some propositions about what is going to happen in the future are not yet either true or false.	B4. The future is as fixed as the past; every proposition must be true or false, and propositions have their truth values eternally (as noted in B1).
A5. Identity through time is *endurance*: numerically the same thing exists at many distinct times.	B5. Identity through time is *perdurance*: a thing that lasts through time is a series of distinct temporal parts or stages, united by some relation other than identity.

In row 1, we have the debate between those who take tense as primitive and those who seek to reduce it to something else (as Smart once did when he suggested that "it will rain" just means "rain occurs later than this utterance".) In row 2, we have the debate between the A theory proper and the B theory proper, which is sometimes too quickly equated with the debate in row 1. (Arguably, tenses are not equivalents of the A- characteristics, but superior substitutes for them.) In row 3, we have the issue that divides *presentists* from *eternalists* – those like AUGUSTINE, who laments that his boyhood is no more, and those like the Tralfamadorians in Vonnegut's *Slaughterhouse Five*, who do not cry at funerals because their departed loved one exists and breathes at an earlier moment. In row 4, we have an issue that goes back to ARISTOTLE's discussion in *De Interpretatione*: must the proposition *the captain will order a sea battle tomorrow* be true or false today, and if so, does that mean the future is in some way fixed or fated? Finally, in row 5 we have the issue (stated in David LEWIS's terms) that divides those who believe in genuine continuants from those who accept an analysis of

identity through time like that of WILLIAMS, who once observed that "each of us proceeds through time only as a fence proceeds across a farm" – that is, by having different parts at different moments or regions (Williams, 1951, p. 463).

As noted, a philosopher who holds a view in one of the columns will tend to hold the other views in that column as well. There is a certain amount of room for mixing and matching, however, and it should not be assumed automatically that the propositions in a given column must go together as a package deal.

Indeed, no one should hold *all* of the propositions in column A, for a little reflection shows that A2 is inconsistent with A3. If presentism is true, there are no things or events that are not present, and thus no items possessed of pastness and futurity. So if A3 is true, A2 is false.

The best combination among A1–A3 for a friend of dynamic time is arguably A1 and A3 without A2. Ironically, this would be an "A theory" without the A-characteristics, so the common name is not well chosen. McTaggart's combination was just the opposite, and this is arguably what led to the demise of time in his philosophy. His argument depends on reducing tense to the A-characteristics, and it also depends on making the eternalist assumption that the earlier and later portions of the B-series are equally real. A presentist could evade the argument by denying that an event is there before it becomes present; rather, the event simply *becomes* – it comes into being and then as quickly passes out of being. Or better yet (since an ontology of things goes better with presentism than an ontology of events), a thing becomes F and then is no longer F.

The issue debated in rows 1 and 2 is sometimes put this way: does time *pass*, or is there simply a huge four-dimensional manifold with time as one of its dimensions? Some philosophers think the passage view may be refuted by asking a simple question: how *fast* does time pass? If the first second of the year 2050 is getting closer to us, there must be a rate at which it is doing this, yet any way of assigning the rate would be nonsensical or absurd. Are the seconds going by at the rate of one second per second? That is no rate at all. One second per hypersecond? That takes the first step in a preposterous series of time orders. So time does not pass.

When the argument is formulated that way, it presupposes a substantival theory of time – as though there were drops of time passing through an hourglass. Perhaps, then, the argument can be sidestepped by combining belief in dynamic time with a rejection of substantival time. Such is the combination espoused by Arthur Prior, the founder of tense logic. Prior represents tenses with operators, akin to modal operators: "Peter will sneeze" becomes "It will be the case that Peter sneezes", symbolizable as *Fp*, and "Peter sneezed" becomes "It was the case that Peter sneezes", symbolizable as *Pp*. The present tense is the default tense and needs no operator. With this apparatus, it is possible to articulate many propositions about the structure of time. For example, the density of time may be expressed as $(p)(Fp \rightarrow FFp)$. This formula would not be true if time were discrete, for if there were an immediately next moment and a proposition p true at it but not thereafter, *Fp* would be true and *FFp* false. Prior denies that time is a literal object, "a sort of snake which either eats its tail or doesn't, either has ends or doesn't, either is made of separate segments or isn't;" rather, these issues can be formulated using propositional variables and tense operators in a way that makes no reference to time or its parts (Prior, 1968, p. 189).

Returning now to the question of time's passage, Prior suggests that the metaphor can be cashed out in tense logic as follows: there are true instances of the schema $Pp \,\&\, \sim p$ – it was the case that p, but is not now the case that p. When the matter is put that way, it is no longer obvious how awkward questions about the rate of time's passage are to be formulated.

PROBLEMS FOR PRESENTISM

Presentism is easily misunderstood. Presentists are not holocaust deniers; their insistence that nothing past exists is

compatible with their affirming truths about what happened using tense operators. Nonetheless, presentism is not without its problems. Are there not past tense truths about individuals who no longer exist, for example, that Lincoln was wise and wore a beard? But how can there be such truths if Lincoln no longer exists to be a constituent of propositions about him? On this question, Prior bites the bullet and says there are no *singular* truths about objects that no longer exist, but only *general* truths – it was once the case that there was a man who was President during a civil war, etc., and who wore a beard. Other presentists find some presently existing entity for past-tense truths to be about – for example, the HAECCEITY *being Lincoln*, a property that exists even if Lincoln does not, and which was formerly co-instantiated with the property of being wise.

What some regard as the fatal blow for presentism comes from the Special Theory of Relativity. The theory is often presented as resting on two postulates, the relativity of uniform motion and the constancy of the speed of light. Uniform motion is motion at a constant speed in a constant direction. The first postulate tells us that no experiment can determine that an object is in a state of absolute uniform motion, from which it is often concluded that it makes no sense to ascribe uniform motion. (If two objects are moving uniformly relative to each other, it is as correct to say that one is moving and the other at rest as vice versa.) The second postulate tells us that whether an observer is moving towards or away from a beam of light, the light's speed with respect to the observer will be the same. Einstein showed that when these two postulates are combined, many surprising consequences follow, including the relativity of simultaneity: two events that are simultaneous in one observer's frame of reference may be successive in another's frame, with no way of saying that either frame is uniquely correct.

PUTNAM has offered an argument against presentism based on Special Relativity and two other assumptions. One assumption (which Putnam calls the principle of "no privileged observers") is that what is real for you is real for me, assuming that you are real for me. This may be expressed equivalently as the assumption that the relation of being real-for is transitive:

1. If x is real for y & y is real for z, then x is real for z.

Putnam's other assumption is that in the context of Special Relativity, the presentist's core thesis that x is real iff x is present should be reformulated as "x is real for y iff x is present for y" and the latter in turn as "x is simultaneous with y in the frame of y":

2. Presentism implies: x is real for y iff x is simultaneous with y in the frame of y.

From 1 and 2, it follows that for presentists, the simultaneity relation we have just mentioned is transitive:

3. Presentism implies: if x is simultaneous with y in the frame of y & y is simultaneous with z in the frame of z, then x is simultaneous with z in the frame of z.

According to Special Relativity, however,

4. The relation in 3 (which Putnam calls "simultaneity in the observer's frame") is *not* transitive.

That is because if you pass right by me at a high relative speed, there will be events simultaneous with you in your frame that are not simultaneous with me in my frame, even though at the moment of passing, you are simultaneous with me in my frame. Putnam concludes that presentism is false, and that I should acknowledge as real events belonging to your present even though they do not belong to mine.

If presentists do not wish to accept this conclusion, how should they respond to Putnam's argument? There are three main options. One is to reject the transitivity of the real-for relation, as advocated by Sklar; in effect, this is to make reality itself as relative as simultaneity. A second is to reject Putnam's construal of "x is present for y" as "x is simultaneous with y in the frame of y"; alternative relativistic reconstruals of the present-for relation have been canvassed by Hinchliff and Sider. The third is to question Special Relativity, as has been done by

Prior. This last response may strike some as an audacious denial of physics to make room for metaphysics, but it need not be that. It will probably not have escaped the reader's notice that insofar as Special Relativity says *there is no such thing* as absolute uniform motion – not just that it is undetectable by any experiment – it ventures beyond physics into philosophy. One who questions the theory may be questioning its verificationist auxiliary assumptions rather than anything that physics alone can teach us.

IS TIME TRAVEL POSSIBLE?

This question turns in part on the issues in rows 3, 4, and 5.

The physics of the last century is sometimes thought to imply an answer of *yes*, for two main reasons. First, the Special Theory of Relativity is sometimes thought to imply eternalism, as discussed above, and the eternalist view encourages us to take time travel seriously. If the assassination of JFK is there, several decades prior to us on the time line, why couldn't we go there and witness it? (Conversely, presentism is sometimes thought to rule out time travel, on the ground that if the past and the future are not there, there is literally nowhere to go.) Second, the General Theory of Relativity is now believed to imply the possibility of closed timelike curves, which might be exploited by time travelers. Einstein's field equations enable one to calculate the spacetime structures induced by various configurations of matter, and in 1949, GÖDEL showed that there are possible configurations of matter that would generate closed timelike curves – temporal paths along which an event can precede other events which precede itself. An object part of whose lifeline lay along such a curve could (in a sense) visit its own past. Interestingly, Gödel's own conclusion from his discovery was quite different: he thought *real* time could not violate the irreflexivity of precedence, so he took the possibility of loops in time to show that time is ideal in something like Kant's sense.

If permitted by physics, travel to the past may nonetheless be forbidden by logic or metaphysics. An entrenched axiom is that no one can change the past. If we could travel to the past, why could we not change it, even in paradoxical ways such as by killing one's grandfather or infant self? Science fiction writers sometimes take pains to have their characters leave the past undisturbed; for example, they view dinosaurs from magically suspended walkways so as to leave no footprints. But of course the mere presence of the time traveler as an observer would constitute a change in the past if he had not been there the "first" (and only) time around. Therefore, in consistent time travel tales, the traveler "always" made his visit – the visit does not change the past, but was always part of it. (As Lewis has it, a temporal stage of the traveler was permanently present at the scene. Lewis's stage view explains how it is possible for the traveler to interact with his infant self: such interaction occurs between stages of the same person that are contemporaneous in "external" time but one later than the other in "personal" time.) Because his actions are already woven into the past, a time traveler cannot kill his grandfather or his infant self; in history as it was, grandfather lived and the traveler failed to kill him, if he tried.

This way of preserving the past from change may arouse fears of fatalism. If in fact grandfather lived to sire my father, am I not fated to fail in my attempts to kill him? And if in history as it happened, I emerged from a time machine in 1920 that I enter (entered? will enter?) in 2020, am I not fated to enter the time machine in 2020, or at least at some time? To do otherwise would be to do something at variance with past truth. In reply, some argue that time-travel arguments for fatalism add nothing to more general arguments for fatalism based on applying the law of bivalence to the future, such as the following:

1. It was either true yesterday that I would push the nuclear button tomorrow or true yesterday that I would not.
2. In the former case, I *must* push the button tomorrow
3. In the latter case, I must *not* push it.
4. Either way, only one course is open to me.

A common reply to such Aristotelian worries is that all that follows from the supposition that it was true yesterday that I would push the button tomorrow is that I *will* push it, not that I *must*. It could be maintained similarly that although in 2020 I certainly *will* enter the time machine from which I emerged in 1920, it is not true that I *must*. So a good case can be made that time travel imposes fatalistic constraints on time travelers only if Aristotelian arguments from bivalence impose fatalistic constraints on us all. So which is it, freedom for time travelers or fate for us all? Space and time do not permit an answer to this question here.

See also the A–Z entries on ANTINOMIES; CHANGE; CONTINUANT; FATALISM; PRINCIPLE OF VERIFIABILITY; SMART, J.C.C.; SPACE AND TIME, TEMPORAL PARTS; ZENO OF ELEA.

BIBLIOGRAPHY

Abbott, E.A.: *Flatland* (New York: Dover, 1952; originally published in 1884).

Broad, C.D.: *Scientific Thought* (New York: Harcourt, Brace, 1920).

Hinchliff, M.: "A Defense of Presentism in a Relativistic Setting," *Philosophy of Science* 67, suppl. (2000), S575–86.

Kant, I.: *Critique of Pure Reason*, trans. N. Kemp Smith (New York: St. Martin's Press, 1965; originally published in 1781).

LePoidevin, R. and MacBeath, M., ed.: *The Philosophy of Time* (Oxford: Oxford University Press, 1993).

Lewis, D.: "The Paradoxes of Time Travel," *American Philosophical Quarterly* 13 (1976), 145–52; repr in LePoidevin and MacBeath (1993).

McTaggart, J.: "The Unreality of Time," in *The Nature of Existence*, vol. II (Cambridge: Cambridge University Press, 1927), ch. 33; repr. in LePoidevin and MacBeath (1993).

Markosian, N.: "How Fast Does Time Pass?" *Philosophy and Phenomenological Research* 53 (1993), 829–44.

Putnam, H.: "Time and Physical Geometry," *Journal of Philosophy* 64 (1967), 240–7.

Prior, A.N.: *Past, Present, and Future* (Oxford: Clarendon Press, 1967).

Reichenbach, H.: *The Philosophy of Space and Time* (New York: Dover, 1958).

Sider, T.: *Four-Dimensionalism: An Ontology of Persistence and Time* (Oxford: Oxford University Press, 2001).

Sklar, L.: *Space, Time, and Spacetime* (Berkeley: University of California Press, 1974). Contains discussions of Newton, Leibniz, and Poincaré.

Sklar, L.: "Up and Down, Left and Right, Past and Future," *Noûs* 15 (1981), 111–29; repr. in LePoidevin and MacBeath (1993).

Van Cleve, J.: "If Meinong Is Wrong, Is McTaggart Right?" *Philosophical Topics* 24 (1996), 231–54.

Van Cleve, J. and Frederick, R., ed.: *The Philosophy of Right and Left: Incongruent Counterparts and the Nature of Space* (Dordrecht: Kluwer, 1991).

Willliams, D.C.: "The Myth of Passage," *Journal of Philosophy* 48 (1951), 457–72.

Yourgrau, P.: *Gödel Meets Einstein* (Chicago: Open Court, 1999).

JAMES VAN CLEVE

Substance

I – Introduction

In one metaphysically salient sense of the term "substance", a substance is an individual thing. From a commonsensical perspective, it appears that the extension of "substance" in this sense includes inanimate material objects, e.g., pieces of gold, mountains, and statues, as well as living things, e.g., people, frogs, and trees. (Note that since a *compound substance* is a *unified* whole, its parts must stand in some sufficiently robust unifying relation to one another, e.g., some appropriate causal or functional relation; if there are *simple* (or basic) *substances*, they do not have any detachable parts, *see* PART/WHOLE.) A belief in the existence of such *individual substances* is at core of our "folk ontology". Moreover, various scientific theories seem to be committed to their existence. The

concept of an individual substance figures prominently in ARISTOTLE's seminal work in metaphysics and in much subsequent important work in the field. It is this concept that is the focus of this essay.

Aristotle's term "primary *ousia*" has often been translated as *substance* (or as *primary substance*) a practice which has caused considerable confusion. This translation can be misleading, since although one ordinary meaning of "substance" is an individual thing, e.g., an inanimate material object or living organism, this is *not* what Aristotle *means* by "primary *ousia*". A more accurate and less misleading translation of "primary *ousia*" is *primary being* (or fundamental entity, or basic entity). In the *Categories* Aristotle argued that the primary beings are individual things, e.g., living things, and that essences are secondary beings. However, in the later work, the *Metaphysics*, he changed his view about primary beings, and seems to have concluded that that the primary beings are forms, rather than individual things. In the *Metaphysics*, Aristotle famously conceived of an individual thing as, in some sense, a combination of form and matter (*see* MATTER/FORM). Even if there exists a technical usage of the term "substance" in which it means *primary being*, this is a different meaning than the more ordinary sense, that of *individual thing*.

But, according to *another* ordinary sense of the term "substance", a substance is *a quantity of material stuff of some kind*, e.g., a quantity of gold, iron, oak, or lamb. But it is one thing to say that there exists a quantity of material stuff of some kind, and quite another to say that there exists an individual substance, even if this individual substance is composed of a quantity of stuff of the kind in question. For example, it is one thing to say that Mary has *50 pounds of lamb*, and quite another to say that Mary has *a lamb that weighs 50 pounds*. After all, a lamb necessarily possesses a certain *form* and *unity* which a quantity of lamb need not possess. Furthermore, it seems possible for there to be an individual substance which has no proper parts, e.g., a non-physical soul or a point-particle; yet, the existence of individual things of these sorts does not entail the existence of a quantity of material stuff of some kind.

The existence of individual substances other than inanimate material objects and living organisms is controversial. However, allowing for the possibility of such substances, including non-physical substances, it is extremely plausible that any conceivable substance is either *spatially extended*, *spatially located*, or *living* (in a broad intuitive sense of "living"). For example, *spatially unextended* or *spatially un-located* substances which have *thoughts*, e.g., Cartesian souls, would qualify as living in virtue of their having *mental* life, even if they lack *biological* or *physical* life (*see* SOUL), whereas apparently *immaterial* physical objects such as point-particles and mass-less extended physical objects would have *spatial location* and/or *spatial extension*. (Hence, given the highly plausible assumption that, necessarily, life is either a physical process or a mental one, it is extremely plausible that any conceivable substance either has spatial extension, spatial location, or thought.) According to SPINOZA, there exists *one* and *only one* individual substance, identical with the universe, and this substance is *neither* a physical substance *nor* a Cartesian soul; still, in Spinoza's view, this substance has both *thought* and *spatial extension*.

II – *The Analysis of Substantiality*

In this section, we shall elucidate what we mean by an *analysis* of the concept of an individual substance, and then discuss the important further notion of the degree to which a philosophical analysis is *ontologically neutral* (*see* ANALYSIS).

We begin with what we mean by an analysis or *analytical definition* of a concept or attribute, *F-ness*. Such an analysis provides a set of conditions, *SC*, such that: (i) an item's (x's) satisfying *SC* is logically or metaphysically *necessary* and *sufficient* for x's being *F*, and (ii) necessarily, if x is *F*, then x's being *F* can be *explained by* x's satisfying *SC*. In this sense, it can be said that an analytical definition of *F-ness explicates F-ness*. However, if being *F* is a *part* of being *C*, then x's being *F* cannot be *explained by* x's

satisfying *SC* on pain of vicious circularity. In such a case, the proposed analytical definition of *F-ness* is fatally flawed; e.g., the proposal to explicate what is *just* as what *conforms to just laws* suffers from this sort of flaw. Circularity of this kind is vicious because nothing can be explained by itself. Hence, necessarily, any purported or candidate analytical definition that involves this sort of conceptual circularity fails to satisfy condition (ii) for being an analytical definition, above, and should be rejected.

Applying this schema to substantiality, let *F* be replaced by *substance*. It follows that in order to provide an analysis of being a substance, an analytical definition must provide a set of conditions, *SC*, such that (i) an item's (*x*'s) satisfying *SC* is logically or metaphysically sufficient and necessary for *x*'s being a substance, and (ii) necessarily, if *x* is a substance, then *x*'s being a substance can be explained by *x*'s satisfying *SC*.

A further important feature of philosophical analyses is to degree to which they are *ontologically neutral.* The following *Principle of Ontological Neutrality* clarifies this notion:

> (*PON*) An analysis, *A*, is ontologically neutral with respect to an ontological kind *K* (or to an entity *E*) =df. The adequacy of *A* does not entail either that *K*s exist or that *K*s do not exist (or that *E* exists or that *E* does not exist).

By the ***adequacy*** of an ontological analysis, we mean that the analysis does not conflict with the ***data*** for that analysis. For example, if one were trying to analyze what a ***concrete entity*** is, then one's analysis should imply that what intuitively are concrete entities are concrete, and that what intuitively are not concrete entities are not concrete. (We shall ignore here the more complicated situation that arises when ***no*** analysis can be formulated that is in this sense adequate to the data, so that we have to choose among proposed analyses none of which is entirely adequate.) It follows from *PON* that if in order to be adequate, a given analysis entails, for example, that universals do or do not exist, or that Cartesian souls do or don't exist, or that God does or does not exist, then it is not ontologically neutral with respect to universals, or to Cartesian souls, or to the existence of God. If an alternative analysis does not have these entailments, and so is ontologically neutral with respect to universals, souls, and God, then, to that extent, the second analysis is more ontologically neutral than is the first analysis. Of course, it may be the case that comparisons between competing analyses are not completely straightforward. It may happen, for example, that analysis *A1* is ontologically neutral with respect to *F*s and *G*s, and not with respect to *M*s and *N*s, while analysis *A2* is ontologically neutral with respect to *M*s and *N*s, but not with respect to *F*s and *G*s. Many other permutations are possible. But at least sometimes, we will be able to say that one analysis is more ontologically neutral than another. In any case, one should be aware of the sorts of ontological commitments assumed by any analysis.

It is plausible to say, we believe, that the more ontologically neutral an analysis is, the better; more precisely, that all other things being equal, analyses having a higher degree of compatibility with the existence of entities of various CATEGORIES are to be preferred, so long as the entities in question are not known to be unintelligible, and plausible views about the nature, existence conditions, and interrelationships of entities belonging to those categories are assumed. Why should this be so? Because which kinds of entities, and which entities, actually or possibly exist, is often a matter of philosophical controversy. Witness the eternal debate over the existence of universals between realists and nominalists. Hence, if one can analyze, say, the concept of substance, without thereby being committed either to the existence or non-existence of universals, then that is preferable, other things being equal, to analyzing this concept in such a way as to be committed to the existence or non-existence of universals. This principle about ontological neutrality seems to us just to be a special case of Ockham's Razor (*see* OCKHAM). It also seems to us likely that there are further principles for evaluating the ontological neutrality of philosophical analyses, but we shall not

attempt to provide a complete statement of them in this article.

In section IV, we shall defend a version of an independence analysis of the concept of substance which is ontologically neutral with respect to a large variety of metaphysical entities: absolute and relational space and time, space-time, universals, tropes, sets, numbers, propositions, events, boundaries, privations (*see* SPACE AND TIME; UNIVERSALS; TROPE; CLASS, SET, COLLECTION; PROPOSITION, STATE OF AFFAIRS; EVENT THEORY; BOUNDARY) and, among substances, living organisms, atoms, ARTEFACTS, and so forth. Other contemporary philosophers have offered competing version of an independence analysis of substance, for example, LOWE (2006) and CHISHOLM (1996). Could there be more than one adequate analysis of substance? We see no a priori reason to rule out such a possibility. One measure of acceptability, however, and one that ought not to be ignored, but is often ignored, is the degree to which such competing analyses are ontologically neutral.

III – Historical Views of Substance

The concept of an individual substance, thing, or object has held a very prominent place in the history of metaphysics, perhaps because it holds such a prominent place in our ordinary conceptual scheme.

In this section, we shall survey several important approaches to analyzing the notion of an individual substance. Among substance realists, there are independence, INHERENCE, CHANGE, and SUBSTRATUM theorists. Also important to consider are those who would reduce substances to items belonging to some other ontological category, and those who argue for their elimination altogether.

Aristotle, in the *Categories*, offers this account of substance in terms of change:

> *It seems most distinctive of substance that what is numerically one and the same is able to receive contraries. In no other case could one bring forward anything, numerically one, which is able to receive contraries.* (*Complete Works*, Vol. I, p. 7)

A sympathetic reading of this attempt to analyze substance is that Aristotle is saying that among entities, only individual substances are able to persist through *intrinsic* change. Hence, Aristotle's analysis of substance in terms of change should be understood as follows:

> (*D1*) *x* is a substance =df. *x* is capable of persisting through intrinsic change.

In the *Categories*, Aristotle lists other categories of being, for example, times, places, qualities, RELATIONS, and kinds. Note that it does *not* seem plausible that such entities cannot persist through *relational* change, as Aristotle appears to have noted. For example, at one moment a particular place might be occupied by a body, while at another time not. However, it *does* seem to be the case that entities of these sorts cannot persist through *intrinsic* change, since they cannot undergo intrinsic change (*see* EXTRINSIC/INTRINSIC).

Nevertheless, there seem to be at least two fairly plausible counterexamples to *D1*. The first is an atomic body, that is, a physically indivisible body. Such substances do not seem capable of undergoing intrinsic change – indeed, that was one of the reasons for the first atomists, Democritus and Leucippus, to postulate such beings (*see* ATOMISM; PRESOCRATICS). Current atomic theory also regards its fundamental particles in this way. Thus, if intrinsically unalterable atoms are possible, then *D1* fails to provide a logically necessary condition for something's being a substance.

The second counterexample to *D1* is provided by boundaries. For example, when a rubber ball bounces, its surface changes its shape. Hence, if there are things like surfaces, and surfaces can undergo intrinsic change, then *D1* fails to provide a logically sufficient condition for something's being a substance.

Each of the preceding counterexamples to *D1* involves a kind of entity that Aristotle did not include in his ontology. Hence, Aristotle could reply that there are no such counterexamples. This points out how Aristotle's *D1* is not an ontologically neutral analysis of the concept of substance: it is not compatible with an ontology that allows for

the possible (or actual) existence of either atomic, intrinsically unchangeable bodies, or of boundaries such as surfaces. Especially in the former instance, this seems to be a serious problem for *D1*.

Aristotle provides a second account of substance in the *Categories*:

> *A substance – that which is called a substance most strictly, primarily, and most of all – is that which is neither said of a subject nor in a subject, for example, the individual man or the individual horse.* (*Complete Works*, Vol. I, p. 4)

This account of substance, then, seems to analyze the notion of substance as follows:

> (*D2*) *x* is a substance =df. *x* can be neither said of nor in a subject.

The basic idea behind *D2* is supposed to be that individual things or substances do not stand in certain relations of dependence to other things, while things in other ontological categories do stand in certain dependence relations to (at least) substances. For example, Aristotle thinks that in the proposition, *Socrates is a man*, the kind, Man, is said of Socrates, implying that Man depends in some sense on Socrates. He also thinks that in the proposition, *Socrates is hungry*, the quality, Hunger, is in Socrates, implying that Hunger depends in some sense on Socrates.

One problem for the idea that *D2* establishes that substances possess a unique kind of independence can be seen by looking at the said-of relation. According to Aristotle, what can be said-of substances are kinds (which Aristotle also calls "secondary beings"), that is, the species and genera under which a substance falls. And given his theory of universals, no substance-kind exists unless it is instantiated by one or more substances. Hence, given Aristotle's ontology, the existence of a substance-kind entails the existence of a substance, so that it might be said that substance-kinds depend on substances. On the other hand, no substance can exist unless it instantiates certain substance-kinds, so it also might be said that substances depend on substance-kinds. Thus, it is not at all clear that the asymmetry of the said-of relation, whereby substance-kinds are said-of substances, but not vice versa, establishes the intended asymmetry of dependence that Aristotle has in mind, whereby substance-kinds depend on substances, but not vice versa.

Similar difficulties attend the claim that, because certain beings are "in" substances, such beings asymmetrically depend upon those substances. Furthermore, it is not clear that on any reasonable understanding of the in-relation employed in *D2*, substances cannot be "in" anything. For example, it seems perfectly natural to assert that a particular body is "in" space and time.

Aristotle's attempt to analyze the concept of substance in terms of the said-of-relation and the in-relation seems to have arisen from certain grammatical features of proper names for individual substances. Such terms can function only as subjects in sentences, and never as predicates. That this fact about grammar can be used somehow to analyze the notion of substance while implying that substances are asymmetrically independent of all other categories of being is, however, an error. If substances do enjoy this sort of independence, and it has been a persistent theme in metaphysics that a correct analysis or understanding of substance will have this implication, then we must seek a different analysis of substantiality than *D2*.

In the later *Metaphysics*, Aristotle defends his hylomorphic account of substance, according to which a substance is a combination of form and matter. On one interpretation, this is just a useful way of distinguishing, in the case of compound bodies, between the structure of the body and its constituent stuff. Such an analysis is level-relative. If Aristotle meant to say that there could be pure (or prime) matter, stuff without form, then this is of questionable coherence. He is also ambivalent about the possibility of the existence of pure form. In any case, Aristotle's HYLOMORPHISM seems incompatible with the possible existence of immaterial souls.

DESCARTES sought a different independence analysis of the concept of substance. For example, at one point he states,

> *The answer is that the notion of substance is just this – that it can exist all by itself, that is without the aid of any other substance.* (*Philosophical Writings*, Vol. II, p. 159)

Hence, Descartes seems to be endorsing the following analysis of the concept of substance:

> (*D3*) *x* is a substance =df. *x* can exist without the aid of any other substance.

This obviously won't do, since to try to analyze the notion of a substance in terms of being capable of existing without the aid of any other substance, is viciously circular. Moreover, *D3* implausibly implies that if God exists, then only God is a substance – for no created substance can exist without the aid of God.

At another point, Descartes avoids the circularity of *D3* with the following statement:

> *By substance, we can understand nothing other than a thing which exists in such a way as to depend on no other thing for its existence.* (*Philosophical Writings*, Vol. II, p. 210)

The implied analysis of the concept of substance is the following:

> (*D4*) *x* is a substance =df. *x* exists and *x* depends on no other entity for its existence.

D4 seems to avoid the circularity of *D3*, but has problems of its own. The main one is that no entity is independent of every other entity. For example, for any entity, *x*, there is a property, *y*, such that *x* has *y* essentially, and thus depends on *y* in the sense of entailing its existence. Another problem is that a compound body, which is a substance, depends on its parts in the same sense. Therefore, *D4* does not appear to provide a logically necessary condition for something's being a substance.

Spinoza is another proponent of an independence theory of substance. His famous definition of substance reads as follows:

> *By substance, I understand that which is in itself and is conceived through itself; in other words, that, the conception of which does not need the conception of another thing from which it must be formed.* (*Ethics and Selected Letters*, p. 31)

Spinoza's definition presents many difficult problems of interpretation, but on the face of it, appears to analyze substance in terms of some sort of *conceptual* independence, with the idea being that what is conceptually independent is also metaphysically independent. Spinoza thought that his definition implied that there was only one substance, Nature, and that this substance exists necessarily. There appear to be at least two serious criticisms of Spinoza's analysis of substance. First, it fails to account for the data that any successful analysis must account for. In this case, Spinoza's analysis implies, contrary to the data, that atoms, living organisms, and finite inanimate compound bodies are not substances – only the universe is. Thus, Spinoza has not succeeded in analyzing the ordinary concept of a substance; rather, he has substituted a radically revisionary notion of his own. (This criticism applies as well to *D3*, above.) Second, it is not clear that even the universe or nature satisfies Spinoza's definition, since in order to conceive of the universe, it seems, one must conceive of one or more of the attributes of nature, e.g., extension.

More recent independence analyses of the concept of substance attempt to conform largely to our intuitions about what entities are substances while capturing a more complex sense in which substances uniquely possess some sort of independence (e.g., Hoffman and Rosenkrantz, 1994; Lowe, 2006).

Some philosophers have tried to analyze the concept of substance in terms of being a subject in which properties inhere. The idea is that there are properties, and then there are things in which properties inhere, namely, substances. For example, Descartes seems to be embracing this theory when he says,

> *Substance. This term applies to every thing in which whatever we perceive immediately resides, as in a subject, or to every thing by means of which whatever we perceive exists.* (*Philosophical Writings*, Vol. II, p. 114)

The inherence theory, however, fails to provide a sufficient condition for something's being a substance, for every entity

is a subject for its properties, and not only substances.

Realizing this, some philosophers have embraced the substratum or BARE PARTICULAR theory of substance, according to which a substance is a concrete individual that has no properties in itself, but instead serves as that in which the properties of ordinary objects inhere in some sense. A ball, on this theory, is not a substance, but rather a whole constituted by a substance/substratum and certain properties. (Alternatively, the ball is a substance, constituted by a substratum and certain properties – the most effective criticism of substratum theories applies to both versions.) Some have attributed this theory to Descartes and/or LOCKE, and among more recent philosophers, the substratum theory has been defended by BERGMANN and, at one point, RUSSELL.

An apparently devastating criticism of any sort of substratum theory is this: it is incoherent to postulate the existence of something that lacks any properties. Nor does the substratum theorist actually refrain from attributing any properties to substrata, since he says that substrata are concrete, that properties subsist or inhere in them, and so forth.

A final type of theory of substance is the BUNDLE THEORY. This, unlike the preceding theories, is a reductionist theory of substance, that is, it implies that substances are aggregates of entities belonging to *another* ontological category. We shall concentrate here on the bundle theory that holds substances to be aggregates of *concrete* attributes or TROPES. Proponents of this sort of theory defend an ontology devoid of both universals and irreducible substances – a simplifying move that they regard as a major strength of the theory. Bundle theorists include Russell at a later stage of his career, AYER, HUME, Herbert Hochberg, and CASTAÑEDA. A recent and novel version of the bundle theory that tries to distinguish between those attributes essential to a substance and those accidental to it has been defended by SIMONS (1994). Bundle theories face several challenges. One is to explicate the relation(s) that is (are) supposed to *unify* the tropes that comprise the bundle. Another is to avoid difficulties that seem to derive from the modal properties of the bundles and from their identity conditions. For example, if a bundle is a (special kind of) collection of tropes, then since collections have their parts essentially, how can a substantial bundle undergo qualitative (or even relational) change?

In addition to debates over the nature or analysis of the concept of an individual substance, metaphysicians have differed over the kinds of individual substances that there are or could be. A familiar controversy of this sort is the one between materialists, dualists, and idealists. Another aspect of this issue is, among material objects, whether or not compound bodies exist, whether or not inanimate compound bodies exist, and whether or not artefacts exist. VAN INWAGEN, for example, has denied the reality of inanimate compound bodies of any sort (while affirming the reality of atomic bodies and organisms), and he has challenged those who assert their existence to provide a satisfactory principle of unity for such objects. Hoffman and Rosenkrantz attempt to do so both for inanimate compound bodies and organisms, though not for artefacts (Hoffman and Rosenkrantz, 1997). Lowe (2006) and Thomasson (2007), on the other hand, defend the view that artefacts, understood as genuine substances, belong in our ontology.

IV – An Analysis of Substantiality

All individual substances belong to the ontological category of Substance. In a broad sense, ontological categories are the more general kinds of entities which (for all we know) could exist. Examples of such categories and sorts of entities which might belong to them are the following: Place (e.g., a volume of space), Time (e.g., an instant), Event (e.g., a process), Trope (in the sense of a concrete "quality", e.g., the particular wisdom of Socrates), Boundary (e.g., a surface), Privation (a concrete entity such as a hole, gap, or shadow), and Collection (in the sense of an arbitrary sum of any concrete entities, e.g., the Moon + the Empire State

Building + Mount Everest). The foregoing examples of categories are species of Concrete Entity. On the other hand, examples of categories which are species of non-concrete or Abstract Entity are Property (e.g., Wisdom), Relation (e.g., Between-ness), Proposition (e.g., that 2 + 2 = 4), Set (e.g., { }), and Number (e.g., 7). Intuitively, the foregoing species of concrete and abstract entities are *peers* in the sense that all of them are at *the same level of generality*. We call this level of generality *Level C*, assuming a hierarchical tree-like taxonomy in which Entity (the *Level A* category) is the *summum* genus, Concrete Entity and Abstract Entity (the *Level B* categories) are the mutually exclusive and exhaustive divisions of this *summum* genus, and the various species of Concrete Entity and Abstract Entity are the Level C categories (*see* CONCRETE/ABSTRACT).

Since the category of Substance is a species of Concrete Entity, it is a Level C category. But how does one acquire the concept of this Level C category? We address this question below.

To begin, according to a plausible empiricist theory of concept formation, one can *acquire* the concept of a *genus* by perceiving instances of one or more *species* of that genus and engaging in a process of abstraction. This plausible empiricist theory entails that one may possess the concepts of certain species before one possesses a concept of the genus that subsumes them – and this is surely true. This process of concept formation involves one's observing certain relevant similarities between the perceived instances of the genus while setting aside inessential dissimilarities between them. In particular, given that *material objects or bodies* are a species of substance, one can acquire the concept of a *substance* by abstracting from one's perceptions of *bodies,* for example, by noticing that they are enduring entities, that they persist through qualitative change, that they exist independently of other entities of the same kind, and so forth, while setting aside inessential observed differences between them such as differences in shape and size.

One reason why people can acquire the concept of a substance via the abstractive process from perceptions of bodies is because people have an intuitive observational concept of a *material object*, an observational concept which does not presuppose the concept of a material *substance*. According to this intuitive observational concept, a material object or body is an entity which has certain perceivable characteristics, including at least certain basic spatial characteristics, which can exist unperceived, and so forth.

Similarly, people have available to them an intuitive concept of a (Cartesian) soul as a non-spatial entity which has certain mental characteristics. This intuitive concept does not presuppose the concept of an immaterial or spiritual *substance*.

By means of the aforementioned process of concept formation, one can see that souls and bodies would belong to a *common level C category* because one can see that souls and bodies resemble one another in ontologically relevant respects. In particular, one can see that, like a body, a soul *can endure, persist through qualitative change, exist independently of other entities of the same kind,* and so forth.

Since an *immaterial* physical object such as a point-particle or a mass-less extended object resembles a body in these ontologically relevant respects, a physical object of this kind also would belong to the level C category in question.

However, people seem to be unable to conceive of anything belonging to this level *C* category other than a PHYSICAL OBJECT (including material objects and immaterial physical objects), a soul, and a Spinozistic substance. This is because we cannot conceive of anything other than a physical object , a soul, and a Spinozistic substance that could endure, persist through qualitative change, exist independently of any other entities of its kind, and so forth.

In what follows, we seek to revive the traditional idea that a substance is an independent or autonomous being. In particular, we argue that the notion of a Level C category can be utilized to analyze the concept of substance in terms of a sort of *ontological independence* which uniquely characterizes any possible substance.

Our proposed analysis of the concept of substance entails that anything that could

belong to the category of Substance must meet certain independence conditions *qua* belonging to that category. In other words, we shall argue that the concept of substance can be analyzed in terms of independence conditions derived from an entity's belonging to a Level C category. Our analysis, *A*, stated below, consists of the conjunction of three independence conditions.

> (*A*) *x* is a substance =df. *x* belongs to a Level C category, *C1*, such that: (i) *C1* could have a *single* instance throughout an interval of time, (ii) *C1*'s instantiation does *not* entail the instantiation of *another* Level C category which satisfies (i), and (iii) it is impossible that something belonging to *C1* has a *part* which belongs to *another* Level C category (other than the categories of Concrete Proper Part and Abstract Proper Part).

In condition (i), by *an interval of time* we mean a non-minimal time. And by *C1*'s having *a single instance throughout an interval of time*, we mean that something instantiates *C1* throughout an interval of time, and that there is no other instance of *C1* in that interval of time.

Although clause (i) of *A* entails that there could be a substance that is independent of any other substance, it does not entail that *every* substance could be independent of any other substance. For instance, clause (i) of *A* is logically consistent with there being a compound substance that is dependent upon its substantial parts. Hence, according to clause (i) of *A*, an entity, *x*, (regardless of whether x is simple or compound), is a substance in virtue of *x*'s belonging to a Level C category which could have a single instance over throughout an interval of time. Clause (i) of *A* characterizes a substance in terms of an independence condition entailed by the instantiability of a certain Level C category.

Clause (ii) of *A* entails that an entity, *x*, is a substance only if *x*'s instantiation of a Level C category is independent of the instantiation of another Level C category which could have a single instance throughout an interval of time. However, although the existence of a substance may entail the existence of entities of another Level C category, for example, properties, in no case is this *other* category such that *it* could have a single instance throughout an interval of time. It follows that the category of Substance satisfies clause (ii) of *A*.

Clause (iii) of *A* entails that an entity, *x*, is a substance only if *x* belongs to a Level C category whose instantiation by an item is independent of any other Level C category (other than two special Level C categories referenced in clause (iii)) being instantiated by a *part* of that item. In general, a part of a physical substance could only be a physical substance or a portion of physical stuff, and a non-physical soul has no parts. Hence, it appears to be impossible for a substance to have a part that belongs to another Level C of the sort in question, for instance, a place, a time, a boundary, an event, a trope, a privation, a property, a relation, a proposition, and so on. Accordingly, the category of Substance seems to satisfy (iii) of *A*.

A is compatible with either of two assumptions. On the first, all individual substances have contingent existence: each substance could fail to exist. On the second assumption, there is a single necessarily existing substance, *G*, such as God, a substance which could not fail to exist. On either of these assumptions, it is possible for there to be a substance, *s*, which exists throughout some interval of time, *t*, without any other substance existing within *t*. On the first assumption there could exist throughout *t* nothing but a single contingent substance. On the second assumption, if *G* exists in time, then there could exist throughout *t* but a single necessary substance; and if *G* exists outside of time, then there could exist throughout *t* but a single contingent substance.

However, it might be objected that if there is an individual substance, then there must be other substances, namely, the (spatial) parts of the individual substance in question. But it is only true that a *compound* substance must be composed of other substances. It is possible for there to be a *simple* substance that has no other substance as a (spatial) part, for instance, a non-spatial soul, a point-particle, an indivisible, spatially

extended, substance, e.g., a Democritean atom. Note that an indivisible, spatially extended substance has spatially extended parts. However, these parts cannot exist independently of the whole of which they are parts. Yet, necessarily, a substance, *s*, is an independent being in this sense: *s* can exist independently of any other contingently existing substance, *s**, unless *s** is a proper part of *s* or *s** helped generate *s*. Since a spatially extended proper part of an indivisible substance fails to satisfy this independence requirement, such a proper part does not qualify as an individual substance. Rather, it is just a *concrete proper part* (of a substance). Such an insubstantial proper part would be an instance of the special Level C category of Concrete Proper Part. The wording in clause (iii) of *A* that excludes the category of Concrete Proper Part from consideration accommodates this possibility of an individual substance that has an entity of another Level C category as a part.

One sort of part in addition to a spatial part is a temporal part. Clearly, it is at least possible for there to be an *enduring substance* that does not have another shorter-lasting substance as a (temporal) part or sub-stage. (In contrast, necessarily, a *temporally extended event* has other shorter-lasting events as TEMPORAL PARTS, STAGES.) Still, arguably, there could be a temporally extended substance that *does have* other shorter-lasting substances as temporal parts, e.g., a four-dimensional physical object in a four-dimensional space–time continuum. But, possibly, there is an enduring, indivisible, physical particle in three-dimensional space (*not* four-dimensional space–time) which *does not have* another shorter-lasting, indivisible, physical particle as a (temporal) part or sub-stage; or possibly, there is an enduring non-spatial soul which does not have another shorter-lasting soul as a (temporal) part or sub-stage. Thus, it *is* possible that throughout an interval of time, *t*, there exists an indivisible substance and no other substance, for example, just one enduring indivisible particle, or just one enduring non-spatial soul.

On the basis of the preceding discussion, we conclude that the category of Substance satisfies the three clauses of *A*. On the other hand, it appear that the categories of Event, Time, Place, Trope, Boundary, Collection, Property, Relation, Proposition, Set, and Number could *not* have a single instance throughout an interval of time. Let us briefly explore the nature of these categories in order to give some indication of how this observation can be supported.

Consider first the categories of Property and Trope. Necessarily, either an abstract property, or a concrete trope, is an entity that stands in lawful logical or causal relations to *others* of its kind. For example, the existence of squareness (or of a particular squareness) entails the existence of straightness (or of a particular straightness). Similar arguments apply to the categories of Relation, Proposition, Set, Number, and so on.

With respect to the category of Place, necessarily, if space exists, then it has an intrinsic structure that it is compatible with the occurrence of motion. This entails that, necessarily, if space exists, then space contains at least *two* places.

In the case of the category of Time, necessarily, if time exists, then it has an intrinsic structure that is compatible with creation, destruction, qualitative change, or relational change. It follows that, necessarily, if time exists, then there are at least *two* times.

With regard to the category of Boundary, necessarily, every boundary is spatial or temporal in character. The existence of a boundary entails the existence of an extended, continuous space or time which contains infinitely many extended places or times. Moreover, necessarily, whatever is *bounded* has a dimension lacked by its *boundary*, e.g., a dimension of thickness, area, length, or duration. Thus, necessarily, if there is one (spatial or temporal) boundary, then there are infinitely many *other* spatial or temporal boundaries.

Consider next the category of Event. Necessarily, an event that occurs over an interval of time is a process. Necessarily, a process involves other sub-processes that are themselves events. Hence, necessarily, if an event occurs over an interval of time, then there is *another* event that occurs within that temporal interval.

Finally, consider the category of concrete entity, Collection. Necessarily, if a collection, c_1, exists throughout an interval of time, t, then c_1 has at least two parts, x and y, both of which exist throughout t. In that case, it appears that there must be a shorter time, t^*, which is a sub-time of t and which is a part of another collection, c_2, for example, a shorter-lasting collection either composed of t^* and x, or composed of t^* and y. Hence, necessarily, if a collection exists throughout an interval of time, then it appears that there is another collection which exists within that interval of time.

This suggests that the category of Collection fails to satisfy clause (i) of *A*. However, *A* also implies that collections are not substances in virtue of their failure to satisfy clause (iii), a clause that requires that it is impossible for an entity of a Level C category has as a part an entity of another Level C category (with the exception of two special categories which are irrelevant here). After all, something that belongs to a collection is a *part* of that collection, and it is evidently *possible* for something that belongs to a collection to be an entity of a Level C category *other than* the category of Collection, e.g., an entity such as a substance, an event, or a place.

In sum, it appears that there could not be just *one* entity of any of the foregoing Level C categories (throughout an interval of time.) Moreover, in each case there is no *other* Level C category which could be instantiated by an entity belonging to the category in question, and which could have a single instance throughout an interval of time. Hence, (clause (i) of) *A* seems to have the desirable consequence that an entity that belongs to any of these categories is *insubstantial*. Clauses (ii) and (iii) of *A* enable this proposed analysis to deal with insubstantial entities of various other kinds.

For example, suppose for the sake of argument that a purple after-image is an insubstantial entity of the irreducible category Sense-Datum. On this supposition, a sense-datum is not an event, a property, a trope, a boundary, and so on. If so, then an after-image belongs to the Level C category of Sense-Datum. But the instantiation of this category entails the instantiation of *another* Level C category that satisfies clause (i) of *A*, namely, the category of substance. After all, there cannot be a sense-datum unless there is a perceiving substance. It follows that the category of Sense-Datum does not satisfy clause (ii) of *A*. Moreover, there is no *other* Level C category which satisfies *A* and which could be instantiated by a sense-datum. Thus, clause (ii) of *A* has the desirable implication that a sense-datum is an insubstantial entity (*see* SENSA).

Finally, consider the Level C category of Privation. In this context, by a privation we mean a concrete entity which is an absence or lack of one or more concrete entities, and which is wholly extended between two or more bounding concrete entities, or else wholly extended between two more bounding parts of a single concrete entity. A privation in this sense is an insubstantial concrete entity. (So, a negative abstract entity, e.g., the proposition that there are no centaurs, does not qualify as a privation in the relevant sense.)

It seems that the category of Privation satisfies clause (i) of *A*. Consider, for example, the possibility of there being nothing but two temporally separated flashes and the period of darkness, *d*, between them. We may assume that in this possible situation *d* is the *only* privation throughout the interval of time in question.

On the other hand, it can be argued that the category of Privation fails to satisfy clause (ii) of *A* for the following two reasons. First, the category of Substance satisfies clause (i) of *A* and this Level C category is other than the category of Privation. Second, necessarily, if there is a privation, then there is a substance, e.g., a substance which flashes, a substance which is perforated, a substance which is shadowed or which casts a shadow, and so on; though, clearly, there could be a (basic) substance without there being a privation.

Still, some have claimed that there could be a flash without there being a substance that flashes, and thus it is controversial whether the existence of a privation requires the existence of a substance. Fortunately, *A* is neutral with respect to this controversy,

since, in any event, *clause (iii)* of *A* entails that privations are not substances. To see this, note that privation, *d*, has as *parts* certain (lightless) periods of time within *d*. These parts belong to the category of Time, a Level C category *other than* the category of Privation. It follows that the category of Privation fails to satisfy clause (iii) of *A*. In addition, there is no other Level C category which satisfies *A* and which could be instantiated by a privation. Hence, clause (iii) of *A* has the desired consequence that a privation is *not* a substantial entity.

It appears that *A* provides a logically necessary and sufficient analysis of the concept of substance in terms of a kind of ontological independence. In the light of the foregoing discussion, it also appears that this analysis is ontologically neutral to a high degree, that is, compatible to a high degree with the existence of entities belonging to various intelligible categories, given plausible views about the nature, existence conditions, and interrelationships of entities belonging to those categories.

See also the A–Z entry on SUBSTANCE.

BIBLIOGRAPHY

Aristotle: *The Complete Works of Aristotle: The Revised Oxford Translation*, ed. Jonathan Barnes (Princeton, NJ: Princeton University Press, 1984).

Bergmann, Gustav: *Realism* (Madison, WI: University of Wisconsin, 1967).

Campbell, Keith: *Abstract Particulars* (Oxford: Blackwell, 1990).

Chisholm, Roderick: *A Realistic Theory of Categories* (Cambridge and New York: Cambridge University Press, 1996).

Descartes, René: *The Philosophical Writings of Descartes*, trans. John Cottingham, Robert Stoothoff, and Dugald Murdoch (Cambridge: Cambridge University Press, 1984).

Hoffman, Joshua and Rosenkrantz, Gary S.: "How to Analyze Substance: A Reply to Schneider," *Ratio* 20 (2007), 130–41.

Hoffman, Joshua and Rosenkrantz, Gary S.: *Substance Among Other Categories* (Cambridge and New York: Cambridge University Press, 1994).

Hoffman, Joshua and Rosenkrantz, Gary S.: *Substance: Its Nature and Existence* (London and New York: Routledge, 1997).

Locke, John: *An Essay Concerning Human Understanding*, ed. Roger Woolhouse (London: Penguin Books, 1997).

Loux, Michael: *Metaphysics: A Contemporary Introduction*, 3rd edn. (London and New York: Routledge, 2006).

Loux, Michael: *Substance and Attribute* (Dordrecht: Reidel, 1978).

Lowe, Jonathan: *The Four-Category Ontology: A Metaphysical Foundation for Natura Science* (Oxford and New York: Oxford University Press, 2006).

Simons, Peter: "Farewell to Substance: A Differentiated Leave-taking," *Ratio* N.S. XI (1998), 235–52.

Simons, Peter: "Particulars in Particular Clothing: Three Trope Theories of Substance," *Philosophy and Phenomenological Research* 54 (1994), 553–76.

Spinoza: *The Ethics and Selected Letters*, trans. Samuel Shirley, ed. Seymour Feldman (Indianapolis, IN: Hackett Publishing Company, 1982).

Thomasson, Amie L.: *Ordinary Objects* (Oxford and New York: Oxford University Press, 2007).

van Inwagen, Peter: *Material Beings* (Ithaca, NY and London: Cornell University Press, 1990).

JOSHUA HOFFMAN AND GARY S. ROSENKRANTZ