Against *S5*

Impossible Worlds in the Logic of What Might Have Been

Nathan Salmón

University of California, Santa Barbara

CONTENTS

it, is raised by Chandler himself in another paper (1976). I shall call our variant of this problem *the Four Worlds Paradox.*[11]

28.1. The Argument

In order to make out this paradox we must assume that some concrete object of a certain sort (table, ship, etc.) is such that it could have originated from matter which differs only in part from the actual original matter. That is, we need to assume that some concrete physical object of a certain sort, say, some ship, could have originated with slightly different parts, as long as some, or perhaps most, of the parts are the same. We do not have to specify exactly how much of the actual original material can be different before one gets a different ship. We need only admit that *some* difference, even if only slight, is allowable. If Kripke's type of essentialism is correct, then no one ship could have originated from *entirely* different matter. It seems then, given our assumption, that there must be some *threshold,* some point at which one more change from the actual original matter must result in a different ship altogether.[12] The difficulty encountered by cross-world identification principles like (V′) and (V″) is easily exhibited in an example which is completely representative of the general case. Consider a possible world w_1 in which a ship a consists of exactly 100 planks of wood. Suppose for the sake of argument that any ship of this particular plan and structure is such that it *could have* originated from a different set of planks as long as 98% of them are the same, and only 2% are different, but that a change of 3% or more in the original material must yield a distinct ship. That is, suppose that the threshold point for allowable variation

---

[11]See Chandler, 1976, p. 108. Chandler credits Robert Stalnaker in connection with the discovery of the particular problem he considers. A similar problem is also noticed by Roderick Chisholm (1973, pp. 584–586, and 1976, appendix B, pp. 148–149). The version of the problem that we shall consider (first presented in Salmon, 1979b) involves a special difficulty not present in Chandler's problem, namely that it is not settled by hypothesis that the two qualitatively indiscernible worlds differ in their accessibility relations. This feature of the problem will be discussed in Section 28.3. Graeme Forbes brought it to my attention that the general problem qualifies as a philosophical paradox.

[12]The assumption of a sharp cutoff point between what the ship definitely could and what it definitely could not have been made from is not necessary to the paradox, thought it greatly facilitates the exposition. Weakening of this assumption is discussed in Section 28.4.

in original matter is 2%. It is important to notice here that we do not assume that in *any* possible world, any ship constructed from 98% or more of the same matter as ship a in w_1 and only 2% or less different matter is a itself. This assumption is somewhat stronger than the analogue of (V″) for ships. We make only the weaker assumption that there are *some* possible worlds in which ship a is constructed from matter as much as 2% different from that in w_1. At present, we leave open the question of whether there might also be other possible worlds in which a different ship, distinct from a, is made from 98% of the same matter as a in w_1. What we shall show, on our assumptions, is that there must be other such worlds. Let us call the planks that constitute ship a in w_1 'P_1', 'P_2', and so on, up to 'P_{100}'. Now surely there is a possible world w_2 in which a ship b is constructed according to the very same plan from planks $P_1, P_2, \ldots, P_{97}, P_{101}, P_{102}$, and P_{103}, where P_{101}, P_{102}, and P_{103} are any three planks that are qualitatively identical with P_{98}, P_{99}, and P_{100}, respectively, but do not even overlap with any of ship a's original planks in w_1. Ship b does not have enough planks in w_2 in common with ship a in w_1 to be ship a itself. It must, therefore, be a numerically distinct ship. Now either of these ships a and b could have originated from a different set of planks as long as 98 of them are the same. Thus there is a possible world w_3 in which ship a is constructed according to the same plan from planks $P_1, P_2, \ldots, P_{97}, P_{98}, P_{102}$, and P_{103}, since the first 98 of these planks are the same as those in w_1. But there is also a possible world w_4 in which ship b is also constructed according to the very same plan from the very same planks, since all but one of them, namely plank P_{98}, are the same as those in w_2. We may diagram the situation as in figure 4.

$$w_1 \qquad\qquad\qquad w_2$$
$$\langle\langle P_1, P_2, \ldots, P_{97}, P_{98}, P_{99}, P_{100}\rangle\rangle \qquad \langle\langle P_1, P_2, \ldots, P_{97}, P_{101}, P_{102}, P_{103}\rangle\rangle$$
$$a \qquad\qquad \neq \qquad\qquad b$$
$$\| \qquad\qquad\qquad\qquad \|$$
$$a \qquad\qquad \neq \qquad\qquad b$$
$$\langle\langle P_1, P_2, \ldots, P_{97}, P_{98}, P_{102}, P_{103}\rangle\rangle \qquad \langle\langle P_1, P_2, \ldots, P_{97}, P_{98}, P_{102}, P_{103}\rangle\rangle$$
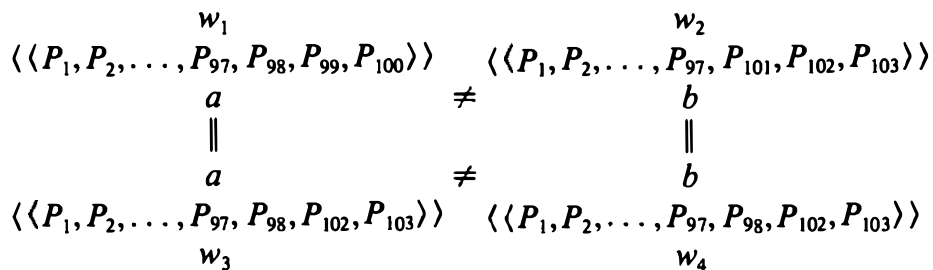$$w_3 \qquad\qquad\qquad w_4$$

FIGURE 4

Hence we have two ships, a and b, in two different worlds, w_3 and w_4, such that both are constructed in their respective worlds (by the very same artisan in the very same place at the very same time) from

the very same planks according to the very same plan; nevertheless they are distinct entities.

Here, then, we seem to have a situation that gives the lie to cross-world identification principles like (V″) and its analogues. For what the argument seems to show is that ships solely constructed from the very same hunk of matter in different possible worlds, and according to the very same plan, cannot always be identified. This conclusion is very surprising in itself, but that is not all. There is nothing in the argument that requires that the two worlds w_3 and w_4 should differ in any way, purely qualitatively or even in the totality of matter they contain, with its exact configuration through time, atom for atom, quark for quark. Nor does the argument require that the entities constituted by the various portions of matter in each world be different, except in the single case of the two ships a and b. And it is difficult to see why variations extraneous to the ships would be called for in any case, since the two worlds are, as it were, minor variants on the initial two worlds w_1 and w_2, and these two worlds may be as alike as one pleases, within the constraint that the two ships differ by three qualitatively identical planks. Thus, for all that the argument requires, we seem to have two possible worlds, w_3 and w_4, that are exactly alike purely qualitatively and even in the very matter they contain with its exact configuration through time, differing only gratuitously over the fact of which ship is constituted by a certain hunk of matter. This is quite paradoxical. How can these two ships, having the very same original matter and structure, not be one and the same ship? After all, it would seem that a ship is nothing *over and above* its parts put together in a certain way, and these two ships, and even the entire possible worlds in which they reside, do not differ in any way qualitatively or structurally. Nevertheless, the correct conclusion seems to be that they differ in their haecceities; the first ship is *this* ship, the second ship is *that* ship, they are different ships, and that is all there is to it. As unpalatable as this may sound, the conclusion seems to follow if we assume that an object might have originated with *some* different parts, but not all. This is a paradox.[13]

28.2. One Solution

I have found that a common reaction to the Four Worlds Paradox is to resist its conclusion by holding that the worlds w_3 and w_4 are really

---

[13]A more precise formulation of the argument for the general case is provided in Salmon, 1979b, pp. 724–725, n. 22.

one and the same possible world, and that we are simply calling a single ship by two names, 'a' and 'b'. Chandler, for instance, flirts with this idea in his discussion of (a variant of) the problem (1976, p. 108)—although he explicitly recognizes that his account affords an alternative solution, one that I accept. The idea that w_3 and w_4 are really identical can take various forms, and has been embedded within various accounts that make an effort to accommodate the initial assumption that artifacts might have been originally made from slightly different matter. The basic idea of these accounts is as follows. Strictly speaking, it is not true that there is a possible world in which the ship that we have called 'a' in w_1—that very entity—is originally constructed from any matter (planks) other than the very same matter (planks) from which it is originally constructed in w_1, even if the difference is ever so slight. In the strict and absolute (haecceitist) sense of 'identical', it is not true that the ship in w_3 that we have called 'a' is identical with the original ship a in w_1; it is not the case that ship a of w_3 is numerically one and the very same thing as the ship a of w_1. Rather, the two ships are, as it were, near misses, not quite numerically identical, but the next best thing. They are proxies or "counterparts" of one another. It is only in virtue of this special connection between the two ships—which is not a strict and genuine identity—that we call the ships by the same name; our doing so should not be construed as a strict identification. We should not say that ship a of w_1 *is* ship a of w_3—or if we do, we must not use the 'is' of strict *numerical* identity but an 'is' with some looser sense (the 'is' of counterparthood?). Similarly for the two ships called 'b' in w_2 and w_4.

This general account may or may not be coupled with a sweeping anti-haecceitism with respect to complex physical objects, restricting the applicability of genuine numerical (haecceitist) identity to more basic entities, e.g., planks of wood, or atoms, etc., or perhaps hunks of matter. In response to the sort of problem posed by the Four Worlds Paradox, Roderick Chisholm, Graeme Forbes, and Kripke have all proposed or suggested one version or another of the basic theory embodied in this account.[14]

---

[14]See Chisholm 1973, 1975, and 1976, appendix B, pp. 145–158. Strictly speaking, Chisholm is concerned primarily with wholes and their constituent parts, though he would probably wish to extend the view to wholes and their constitutive matter. (In this connection see footnote 2 above. See also Hirsch, 1976, p. 55, n. 9.) Kripke (1972a, p. 51, n. 18) suggests a version of the counterpart theory (also concerned primarily with the part-whole relation) in response to problems in this same cluster,

Whatever particular form this account takes, it flies in the face of a straightforward, literal construal of the initial, plausible assumption that some artifact is such that it might have been originally made of slightly different matter. Suppose I point to a particular ship or table and say "Of course, *this very artifact* might have been originally made of mostly the very same wood, with only an atom or two different here and there." I make no reference—explicit or implicit, literal or metaphorical, direct or allusive—to any ship or table other than the one I am pointing to. I am saying something about it—that very artifact—and not about any of a range of imposters. Things might be different if I were to use a nonrigid definite description like 'the unique artifact most like this one'. But I do not, and I do not whisper it either. I use a demonstrative, an obstinately rigid designator, and I point to a particular actual artifact to say something—something true—about it and it alone.[15] I do not mean "There might have been an artifact here, not really this very artifact, but a reasonable facsimile made from mostly the same

---

although he does not explicitly advocate the theory, and says that the particular version he considers seems utopian in its assumption that there is some ultimate or basic kind of entity out of which complex physical objects are made and for which cross-world identity is unproblematic. (More on Kripke in Section 28.4 below.) This version of the theory, or something closely akin to it, is defended and worked out in detail by Forbes (1981b), where the theory is explicitly put forward as a solution to the Four Worlds Paradox. See also Wiggins, 1980, pp. 97–98. Chisholm (1975, pp. 92, 96–97) traces the essentials of the theory to Bishop Butler, David Hume, and Thomas Reid. It goes without saying that David Lewis—the founder of contemporary counterpart theory—advocates a version of the theory sketched in the text, though Lewis, of course, goes much further than the other authors mentioned here. See, e.g., Lewis, 1968. It is not a necessary part of the theory sketched in the text that a ship exists in only one possible world. Nor does the theory require that a ship can exist in distinct possible worlds. It requires only that a ship have the same original matter wherever it exists. Further requirements yield proper extensions of the theory.

[15]Cf. Kripke, 1972a, pp. 43–47. Kripke recognizes the tension between these considerations and his footnote 18, and he attempts a sort of reconciliation at pages 50–53. I believe that the tension is greater than Kripke acknowledges here, as I will try to show in Section 28.4 below. Kripke writes: "Although we can try to describe the world in terms of molecules, there is no impropriety in describing it in terms of grosser entities: the statement that *this table* might have been placed in another room is perfectly proper, in and of itself. We *need* not use the description in terms of molecules, or even grosser parts of the table, though we *may*" (p. 51). Similarly the statement that *this table* might have been made from mostly, but not exactly, the same matter is perfectly proper, in and of itself, and as Kripke seems to argue (p. 45, n. 13), it is a statement about the very table demonstrated, not about various of its imposter "counterparts."

234

wood that this was actually originally made of." If that were what I meant, that is what I would have said.[16] But if for some reason I am wrong about this, and what I said does mean what the counterpart theorist says it means, then let him or her tell me how to say what I meant to say, and I will construct a new Four Worlds Paradox. Better yet, interpret the above formulation of the argument for the paradox in the way that it was intended; then the idea that $w_3 = w_4$ is quite definitely mistaken.

[16]Hazen (1979, pp. 320–322) objects to arguments of this sort on the grounds that they confuse pronouncements of ordinary subjunctive discourse, pronouncements about which we may have firm modal intuitions, with the counterpart theory's proposed analysis into possible world discourse using the technical notion of *counterpart,* where this analysis "is not a sentence that we, qua speakers of our particular natural language, are entitled to have intuitions about." I do not find Hazen's objections convincing, though there is not the space here to develop a full response. In any case, taken as an objection to the argument given in the text, it misses the point: Intuitively the modal operator sentences 'It might have been the case that: *this* be made from hunk H' and 'It might have been the case that: something (most and sufficiently) resembling *this* in certain (such-and-such) respects be made from something (most and sufficiently) resembling H in certain (such-and-such) respects' differ in their truth-conditions. The latter assertion is considerably weaker than the former, if not entirely independent of it. (The latter assertion can only be spelled out further insofar as the relevant notion of *counterpart* is fully and adequately explained in terms of resemblance, and counterpart theorists do not always agree in their explanations. Perhaps it would be better to replace the assertion with one explicitly involving the relevant notion of *counterpart:* 'It might have been the case that: a counterpart of *this* be made from a counterpart of H'. However the relevant notion of *counterpart* is plausibly explained short of strict identity, this assertion is, intuitively, considerably weaker than the first assertion, if not entirely independent of it. Moreover, I know of no plausible grounds for prohibiting the *counterpart* notion from modal operator discourse.) Yet the counterpart theory assigns both (all) assertions the same truth-conditions that standard possible-world semantics, supplemented if need be with the relevant notion of *counterpart,* assigns only to the latter. (I assume here Lewis's principle that any possible object is its own sole counterpart in its own world, or minimally that all of a possible object's existing counterparts in a given world are counterparts in that world of anything that the object is itself a counterpart of in that world. Forbes's original treatment in 1981b preserves the latter principle, and it is difficult to imagine a philosophical motivation for counterpart theory which plausibly rejects the principle on independent grounds.) Counterpart theory fails to distinguish the truth-conditions of the two assertions, and thereby misrepresents the full force of the first assertion. The intended force of the first assertion is something much stronger than (or at any rate, different from) the counterpart theory's surrogate, and something that the theory must reject as false. Contra Hazen, this makes the theory an extreme form of essentialism, despite the deceptive fact that the theory can accommodate the first modal operator sentence, which ostensibly belies extreme essentialism.

The fact is that the counterpart theory sketched above is, at bottom, just a particularly inflexible brand of essentialism. By denying that there is a possible state of affairs in which the very ship a from w_1 is made from ever so slightly different matter, the counterpart theorist holds, even if only tacitly, that in a strict sense it is absolutely impossible for the very ship a, rather than some other ship, to have been made from matter even only an atom different. The counterpart theorist can mouth the words 'Ship a might have been made from slightly different matter', but any such pronouncement by the counterpart theorist in modal operator discourse is a verbal camouflage that merely postpones the inevitable. What matters is what the counterpart theorist *means* by these words, and more importantly, what is *not meant* by these words. (Cf. Plantinga, 1974, pp. 114–119.) To use an analogy suggested by Kripke, the phenomenalist can *say* 'There is a table in the room', but this does not alter the fact that he or she believes in neither the table nor the room. According to the counterpart theory, *strictly speaking,* if in another possible state of affairs we are ever to have the one and only, one and the very same ship a from w_1—that very entity and no other—it must be made in that state of affairs from the *very same* matter, atom for atom, quark for quark, no exceptions. Otherwise, *strictly speaking,* we get a new and different ship. The theory does not *identify* the worlds w_3 and w_4 so much as it denies their existence (or their possibility). It replaces the two of them with a single world containing neither ship a nor ship b, but some new ship.[17] The theory

---

[17]Forbes (1980, pp. 359–360, and 1981a, p. 81) discusses a close analogue of the problem posed by the Four Worlds Paradox for the special case of organisms and the gametes from which they sprang. In these papers he urges certain addition arguments for essentialism and puts forward a variant of the Four Worlds problem as part of a *reductio ad absurdum* argument which, if it worked, could be generalized (as Forbes apparently intends) to establish the particularly inflexible brand of essentialism concerning artifacts inherent in the counterpart theory. To couch his discussion in terms of our example, Forbes implicitly identifies (I would say *confuses*) the two worlds w_3 and w_4 and concludes that the initial assumption that there is such a world, where the very individual from another world has a partly different original composition, has the "unacceptable consequence" that the individual (in our case, ship) in the third world (the schizophrenic w_3/w_4 world) is the same thing as a and distinct from b from the point of view of w_1, but is just the reverse from the point of view of w_2, thus making the "identity" (haecceity) of the individual (ship) in the third world "change according as w_1 or w_2 is supposed to be actual" (1980). But this, whatever it means, is not a consequence of the view in question. Nor does the view contradict any version of the necessity of identity, as Forbes claims (1981a). Forbes's discussion suffers from a

then does its best to accommodate our plausible modal pronouncements concerning the two original ships *a* and *b* without reintroducing the paradox by reinterpreting these pronouncements in terms of the single replacement world. But the reinterpretation is a misinterpretation, and it is uncharitable to those of us who, on reflection and in full command of our faculty of modal intuition, reject any inflexible brand of essentialism that cannot accommodate a possible scenario in which the very ship *a* from w_1 is made from ever so slightly different matter. Still, the argument of the Four Worlds Paradox appears to show that this inflexible brand of essentialism is the price one must pay for cross-world identification principles like (V″), in conjunction with the plausible view that the very same artifact could not have been originally made from entirely different matter.

In fact, the temptation to identity w_3 and w_4, despite the contortions that this identification requires, might be traced in part to the organizational role that identification principles like (V″) play in our way of looking at things. (See Section 26.) It might be traced even further to a possibly more fundamental and more general principle of modal thought. This is the "reductionist" principle that physical objects are "nothing over and above" their matter and structure, in the sense that a complete accounting of what matter there is in a genuinely possible world, with its causal interconnections and exact configuration through time, atom for atom, quark for quark, must completely and uniquely determine whatever physical facts there are about each of the physical objects such as tables and ships present in the world, including such facts as that a particular hunk of matter a' constitutes a particular ship *a* at time *t*. This principle would require that any two genuinely possible worlds exactly alike at the level of

---

flaw analogous to the one we uncovered in Chandler's discussion of his two Ship of Theseus worlds: In deriving the unacceptable "consequence," Forbes illegitimately uses a single expression ('O_2', and the same is true of his use of 'O_3') as a name for what, on the view in question, are distinct individuals originating from the same matter in different (and perhaps mutually inaccessible) worlds. This makes the view seem incoherent when it is not. In fact the view is perfectly tenable, and in the case of ships and tables it is even true. (Caution: Comparison of Forbes's *reductio* argument and our formulation of the Four Worlds Paradox is complicated by the fact that Forbes calls the worlds playing the roles of our w_1 and w_2, 'w_1' and 'w_4', respectively. The situation is complicated still further by the fact that Forbes considers not one but two additional worlds, which he calls 'w_2' and 'w_3', each of which involves a conflation of the distinct roles played by w_3 and w_4 in our example.)

matter and structure must also be exactly alike at least in all their physical-object facts.[18] Assuming as a plausible principle the identity of factually indiscernible possible worlds (worlds in which the very same facts obtain), and given that the worlds w_3 and w_4 are exactly alike in all their non-purely-physical facts, we seem to have that $w_3 = w_4$. If the "reductionist" principle is extended also to self-contained *portions* of full-blown possible worlds, so that a complete accounting of all the causal relations and the exact configuration of a *particular* hunk of matter through time completely determines all the physical facts about the object, if any, so constituted, then we obtain (V") and its analogues.

28.3. A Better Solution

Despite appearances, the conclusion of the argument of the Four Worlds Paradox does not conflict with this reductionist principle, nor does it conflict with cross-world identification principles like (V"). To see this we must turn to Chandler's own example (1976). Chandler's main concern is to argue that the accessibility relation between possible worlds (w is *possible relative to w'*) is not transitive. Considerations similar to those that generate the Four Worlds Paradox seem to yield this result. Given the existence of possible world w_3 in the example, there is by hypothesis yet another world w_5, possible relative to w_3, in which the same ship a is originally constructed by changing one more of its original planks, say P_{98} to P_{101}, so that ship a is now constructed in precisely the same way as the distinct ship b in w_2. But even though w_5 is possible relative to w_3, and w_3 is possible relative to w_1, w_5 is, by hypothesis, not possible relative to w_1. Ship a has exchanged one too many planks. We can put the point as follows. Suppose that w_1 is the actual world, or the way things actually are. Then ship a is in fact constructed from planks P_1 through P_{100}. It is possible for the same ship a to have been constructed from the planks in w_3, since this involves only a change of two planks, but it is impossible for the very same ship a to have been constructed from the planks in w_5, since this would involve a change of three planks from its actual original composition. But, if ship a *had been* originally constructed from the planks in w_3, i.e., if it *had been* constructed with only two different planks, then it would have

---

[18]The principle is explicitly mentioned by Kripke (1972a, p. 50). The principle, or one very much like it, is also explicitly mentioned and endorsed by Forbes (1980, pp. 353–355, and 1981a, p. 79).

been possible for the very ship a to have been originally constructed from the planks in w_5, since this only involves changing one plank from w_3. It *would have been* possible, but it is not *actually* possible! From the point of view of the way things actually are, a scenario or state of affairs in which ship a is originally constructed from the planks in w_5 is not a real possibility concerning ship a. World w_5 is an impossible world—an impossibility as far as w_1 is concerned. World w_3, on the other hand, is a genuine possibility concerning ship a, and if w_3 had been the case, some *new* possibilities concerning the very ship a would arise—possibilities as far as w_3 is concerned. In particular, the scenario or state of affairs w_5 would then become a genuine possibility concerning ship a. Only a narrow-minded form of modal ethnocentrism would deny this. If it is possible in the actual state of affairs w_1 to have altered slightly ship a's origin, then surely it is also possible in the state of affairs w_3 to have altered a's origin *as it is there*. Although w_5 is an impossible state of affairs from the point of view of the actual state of affairs w_1, it is only *contingently* impossible. (Cf. Fine, 1977a, especially p. 139.) It is what might be called *possible in the second degree,* i.e., possibly possible. Even from the point of view of w_1 it might have been a possibility concerning ship a, and indeed it would have been a possibility if only a had had a different plank or two.[19]

In our presentation of the argument for the Four Worlds Paradox, we are given that the worlds w_2 and w_3 are possible relative to the initial world w_1, and that w_4 is possible relative to w_2. But we are not given any reason to suppose that w_4 is also posible relative to either w_1 or w_3. Nor are we given that w_4 is impossible relative to either w_1 or w_3. Thus, what the argument of the Four Worlds Paradox shows is that there are pairs of qualitatively indiscernible worlds w_3 and w_4, each of which contains only one of two distinct artifacts a and b made in their respective worlds from the very same matter, and such that w_3 is possible relative to a given world w_1 (e.g., the actual world) and w_4 is possible relative to an intermediate world w_2 possible relative to w_1. The question remains open whether w_4 is possible or impossible relative to the given world w_1, and whether w_3 and w_4 are possible or impossible relative to each other. This result need not be taken as a threat to cross-world identification principles like (V″) or to the

---

[19]As with the argument of the Four Worlds Paradox, the general argument for intransitivity does not require the assumption of a sharp threshold between what is possible and what is not possible in the construction of a particular artifact. Weakening of this assumption is discussed in the next section.

general "reductionist" principle mentioned in the preceding section, unless some further argument can be given to show that w_4 is also possible relative to w_1 (or that w_3 and w_4 are possible relative to each other).

On the contrary, such principles might be taken as *showing* that w_4 cannot be possible relative to w_1. This would resolve the paradox. Indeed, this is (more or less) the position I wish to take with respect to the paradox.[20] Suppose that there is a world w in which a table is the only table made from the very matter that originally made up an actual table T, and it is made following the same plan, atom for atom. Is this table the same table T? If we are given that w is a way things genuinely might have been, i.e., a world possible relative to the actual world, then the answer seems to be 'yes'. But given that w is a way things could not have been, and is possible in the nth degree for some $n > 1$, i.e., possibly, possibly, ..., possible, who knows? Certainly our belief in (V") *per se* does not yield any continuing inclination to answer our question affirmatively.

Thus principles like (V") are rescued from the Four Worlds Paradox.

28.4. Vagueness and the Paradox

By considering a succession of 51 worlds beginning with w_1, one can eventually construct ship a of entirely different matter than that used in w_1, though on the account just sketched the 51st world in this sequence is a far-removed impossible world from the point of view of w_1, possible only in the 50th degree. This is a sorities construction.[21] Sorites arguments are notorious for playing havoc with the phenomenon of vagueness. In constructing the argument of the Four Worlds Paradox, as well as the argument for intransitivity of modal accessibility, we assumed that there is a sharp threshold or cutoff point between what matter could and what matter could not originally constitute a particular ship a. But this assumption is philosophically unsubtle, and seems too crude and vulgar to be true. Shouldn't we recognize an interval of vagueness between these two extremes, rather than a sharp cutoff point? Any ordinary artifact such as a ship

[20] I am indebted to Penelope Mackie for helping me see my way to this position, though I do not know whether it is her position.

[21] It is very reminiscent of arguments given by Chisholm (1967; 1973, pp. 584–586; and 1976, appendix B, pp. 148–149)—though Chisholm draws an inflexible essentialism as his conclusion rather than the intransitivity of modal accessibility. See also Wilson, 1959.

240

is such that it might have been made from a different hunk of matter as long as it is very nearly the same matter, with only a few tiny replacements here and there. Furthermore, any ship made from a hunk of matter substantially different from the actual matter of an ordinary ship must be a different ship. Between these two extremes it would seem that there is an interval of vagueness, a region of indeterminacy. For hunks of matter in this region—involving a significant amount of overlap with the original ship's actual matter, but also involving a significant amount of new and different matter—it apparently becomes vague or indeterminate (neither true nor false, there is no objective fact of the matter) as to whether a ship made from that very matter in just the same way could be the same ship as the actual ship with which we began. If this is so, any argument that leads to a paradoxical conclusion by ignoring this element of vagueness may be playing havoc with vagueness in just the manner of a sorites argument. Is that what is going on in the Four Worlds Paradox and in Chandler's argument for intransitivity of modal accessibility?

Yes and no. Certainly these considerations of vagueness are relevant to the general problem, but they do not yield a better solution, only something of a better understanding of the general phenomenon. To see this, let us begin by raising a question: Wherein does the vagueness reside? Which terms or concepts involved in the arguments are vague or indeterminate?

Kripke's suggestion of a version of the counterpart theory as part of a possible response to these problems is predicated on the idea that in some instances "the identity relation is vague." He says that

> perhaps, . . . given certain counterfactual vicissitudes in the history of the molecules of a table, T, one may ask whether T would exist, in that situation, or whether a certain bunch of molecules, which in that situation would constitute a table, constitute the very same table T . . . . in concrete cases we may be able to answer whether a certain bunch of molecules would still constitute T, though in some cases the answer may be indeterminate (1972a, pp. 50–51).

In a footnote to this passage, Kripke writes:

> There is some vagueness here. If a chip, or molecule, of a given table had been replaced by another one, we would be content to say that we have the same table. But if too many chips were different,

we would seem to have a different one. . . . Where the identity relation is vague, it may seem intransitive; a [chain] of apparent [identities] may yield an apparent nonidentity. Some sort of 'counterpart' notion . . . may have some utility here. . . . Logicians have not developed a logic of vagueness (n. 18; brackets indicate Kripke's corrections of the original printing).

It is only for cases "where identity is vague" that Kripke proposes that the counterpart theory might be called into active duty, for lack of a logic of vagueness (at least, for lack of a logic of vagueness to Kripke's knowledge circa 1970). But what is such a case? As we noted above, it would seem that for any actual table T there is a region of indeterminacy concerning the table's potential for having had a different origin. For any hunk of matter in this region of indeterminacy, being significantly different from table T's actual matter but also involving just enough overlap, there is no definitive answer to the question of whether a table T' made from that very matter is the same table T or a different table. The sort of cases that concern Kripke, and prompt him to say that identity is vague, are precisely such cases in which we are given a possible world w, with let us suppose a complete accounting of what matter it contains and its exact configuration through time, in which a table T' is constructed with significantly different matter from that of an actual table T, but just enough of the same matter so that there does not appear to be any definitive answer to the question of whether it is the same table.[22]

One must be careful not to be misled by this description of the situation. As described, it is very much like instructing someone to suppose an arbitrary possible world w in which a communist x is elected president of the United States, and then being asked "But who is x, this lousy would-be spy for the KGB?"—questions of the sort that Kripke rightly rejects (e.g., in 1972a, pp. 42–47). There are some worlds in which Angela Davis is elected president as a

---

[22]Derek Parfit (1971) apparently urges a version of the theory of vague or indeterminate identity—one which he regards as entirely uncontroversial for certain sorts of things—as a solution to the problem cases of personal identity. He writes: "No one thinks [that the question of whether a thing x is identical with a thing y must have a true answer] about, say, nations or machines. Our criteria for the identity of these do not cover certain cases. No one thinks that in these cases the questions 'Is it the same nation?' or 'Is it the same machine?' must have answers" (p. 3).

communist, and there are other worlds in which Milton Friedman is elected president as a communist. (If Friedman is essentially noncommunist, then these latter worlds are impossible ones.) There is no answer to the question 'Who is x?', not because identity is vague, but because 'x' is a variable with a range of values. Our description of w did not pin down a particular possible world, with a particular person x, but a *class* of worlds with different persons x. There is an important difference between this sort of case and the table case that interests Kripke. There is no reason to suppose that there is only one person who might have been elected president as a communist, and consequently there is no reason to suppose that we have singled out a unique person x. But in the case of the tables T and T', a believer in (V″) or in the reductionist principle mentioned in Section 28.2 has a reason to believe that there is only one possible table that might have been (the only table) constructed from the relevant matter.[23] On this assumption, we have singled out a unique possible table T', and the question of whether this table T' is the actual table T is meaningful. If the relevant hunk of matter in w differed more extensively from the actual matter of table T, then we could safely say that T' cannot be the same table as T. If the matter in w overlapped more extensively with the actual matter of table T, then we could safely say that T' must be the same table as T. But in the situation described, no such answer is forthcoming; there is no objective fact of the matter as to whether T is T'. So it seems that the 'is' of identity is not defined for this case.

There are serious problems with this way of looking at the situation. Insofar as I understand the idea that identity is sometimes vague, it is provably mistaken. For suppose that there is a pair of entities x and y (e.g., Kripke's tables T and T', or the original ship of Theseus a and the modified ship of Theseus c, times, events, etc.) such that it is vague (neither true nor false, indeterminate, there is no objective fact of the matter) whether they are one and the very same thing. Then this pair $\langle x,y \rangle$ is quite definitely *not* the same pair as $\langle x,x \rangle$, since it is determinately true that x is one and the very same thing as itself. It follows that x and y must be distinct. But then it is not vague whether they are identical or distinct.

---

[23]Kripke's discussion (1972a, pp. 50–51) is in some sense contingent on the assumption of the reductionist principle, though it is unclear to what extent Kripke intends to endorse the principle. See also Chapter One, footnote 41.

This argument is brief, but it is telling.[24] It might be replied that the argument shows no more than that, for any entity x for which there is a y such that the English 'is' of identity is undefined for the pair $\langle x,y \rangle$, the 'is' of identity must also be undefined for the reflexive pair of x and x itself, so that x lies entirely outside of the category of

[24]An argument similar to the one just presented is given in Evans (1978), though the argument had occurred to me independently. It is, I believe, an argument that is straightforwardly demanded by the correct conception of identity as the stark relation between every object and *itself* (as opposed to richer relations like *being constituted by substantially the same matter*).

I have encountered a number of objections to the argument, but none that are convincing. Perhaps the most frequent objection is the idea that if we take vagueness and indeterminacy seriously, it is fallacious to infer that $\langle x,y \rangle \neq \langle x,x \rangle$ from the assumption that it is indeterminate or vague whether the first pair of objects stand in the identity relation, whereas it is fully determinate and settled that the second pair of objects so stand. The objection is usually based on the notion that where a term is applied to objects for which the term's applicability may be vague or indeterminate, classically valid inference patterns are no longer legitimate. But the inference drawn here is from a conjunction consisting of an *assumption*—something we are taking to be determinately the case for the sake of argument—together with something that is quite definitely the case. The inference pattern need only be valid, i.e., truth-preserving. There is nothing more to require of it. Analogously, the term 'bald' may be vague in the sense that there are individuals for whom it is indeterminate whether the term correctly applies, or individuals of whom the term is neither true nor false, and yet the inference pattern $\ulcorner \alpha$ has a full head of hair $\therefore \alpha$ is not bald$\urcorner$ is perfectly valid. The validity of this inference pattern ensures that in order to settle the question of whether it is true, false, or indeterminate that Harry is bald, it is *sufficient* to take note of the fact that Harry's head is very hairy.

For those who may be concerned that the argument against vague identity—which is essentially a metatheoretic argument about the English 'is' of identity—ignores the well-known strictures against conflating object language and metalanguage, it should be noted that the argument could be reformulated entirely within any second language adequate as a metalanguage for English. Alternatively, the formulation of the argument can be left intact, but applied only to an impoverished fragment of English, as object language, which consists only of quantificational idioms and the 'is' of identity. The argument, unlike the suspect arguments of the Liar and the other semantic paradoxes, is a semantical argument about a *nonsemantical* predicate. The crucial assumption is only the modest metatheoretic assumption that the 'is' of identity is determinately true of any object and itself, and determinately false of any pair of distinct objects.

Of course, it might in some cases be "vague" in an epistemological sense (e.g., in principle unknowable) whether a given entity x is identical with a given entity y, but that is a separate matter from the issue before us, and need not entail that there is no objective (albeit unknowable) fact of the matter, or that the 'is' of identity is undefined for the pair $\langle x,y \rangle$. One of the virtues of Kripke's discussion is that he sharply distinguishes metaphysical from epistemological issues, and makes it clear that he is discussing a metaphysical problem about cross-world identity.

meaningful application of 'is'. (Otherwise, by the above argument, x and y would be discernible, hence distinct, and hence the 'is' of identity would be determinately false of them.) But this reply is odious. For the theory of vague identity, and the problems on which the theory is brought to bear, would require that for nearly *any* physical object (ship, table, etc.) x, there will be an object y (in another possible world or at another time) such that it is indeterminate whether $x = y$, and hence, according to this reply, it will also be indeterminate whether $x = x$! More significantly, the reply is ineffective. For suppose that there are entities x for which there is a y such that the English 'is' of identity is neither determinately true nor determinately false of the pair $\langle x,y \rangle$; hence according to the reply, the 'is' of identity is also undefined for the reflexive pair $\langle x,x \rangle$. Now consider the following prophetic argument:

> I shall introduce an artificial relation called 'schmidentity' (not a word of English) which I now stipulate to hold between [*every*] object and itself. Now then . . . the same problems will hold for this [predicate] as were thought in the case of the original [predicate]. . . . If anyone thinks about this seriously, I think that he will see that therefore probably his original account of [vague] identity was not necessary, and probably not possible, for the problems it was originally meant to solve, and that therefore it should be dropped, and identity should just be taken to be the relation between [*anything*] and itself. This sort of device can be used for a number of philosophical problems (Kripke, 1972a, p. 108, taken with gross liberties and out of context).

The assertion that identity is vague is only so much barking up the wrong tree. The 'is' of identity is not vague in the way that an inexact or fuzzy term like 'bald' (in the sense of 'nearly absolutely bald') is vague. Insofar as principles like (V″) or the reductionist principle mentioned above lead to the consequence that identity is vague, we should refuse the principle rather than swallow its consequence. In the case that concerns us, we appear to have a specific and definite table T' that cannot properly be said either to be the same or to be a different table from the table T. But T is an *actual* table, and surely the 'is' of identity is defined for it, determinately true of the pair $\langle T,T \rangle$. Insofar as there is a definite possible table T' such that, for whatever reason, the 'is' of identity is not defined for the pair $\langle T,T' \rangle$, T' simply is not one and the very same thing as T, it is not "schmidentical" with T. There is an objective fact of the matter. On

245

the other hand, if our complete description of the configuration of matter in *w* does not single out a unique and definite possible table *T'*, but only a class of worlds with various tables *T'*, then the question 'Is *T'* identical with *T*?' is as illegitimate as the question 'Who is this would-be spy for the KGB?' In either case, vague identity is not the problem, nor the solution.

Still, the general phenomenon of vagueness seems to play a role in the Four Worlds Paradox, and in Chandler's argument for intransitivity of modal accessibility. For it does seem that, in at least most cases, there simply is no sharp threshold or cutoff point, as the two arguments assume, between what variation in the original matter is possible and what variation is impossible in the construction of a given ship or table. It seems only crude to deny that for any ordinary table *T* there must be a region of indeterminacy such that it is vague or indeterminate whether it is possible for *T* to have been constructed from a certain hunk of matter involving some overlap and some nonoverlap with the actual matter of *T*. For hunks of matter in the region of indeterminacy, there does seem to be no objective fact of the matter as to whether it is possible for the actual table *T* to have been constructed from that very matter, even though it is definitely possible for that very matter to have been made into a table *T'*. Wherein does the vagueness or indeterminacy reside if not in the identity or nonidentity of *T* and *T'*?

One place to look is to the predicate 'constitutes', or the 'is' of constitution. Perhaps in a possible world *w* in which a hunk of matter in the region of indeterminacy for table *T* is made into a table, what is indeterminate is whether that very matter constitutes the very table *T* in *w*. But take notice of what we have just said. For if the relevant matter is formed into a table in *w*, then there is a possible table *T'* in *w* that is constituted in *w* by the relevant matter. Now *T'* cannot be the same table as *T*, for we are given that it is determinately true, and an objective fact, that the relevant matter constitutes table *T'* in *w*, but we are supposing that it is indeterminate, and that there is no objective fact, as to whether that very matter constitutes table *T* in *w*. For this to be the case, it would seem that it must be in some sense a "live option," not ruled out by the objective facts, that the relevant matter should originally constitute both tables *T* and *T'* simultaneously, even though the tables are distinct. If this were ruled out by the objective facts of the matter, then contrary to our hypothesis, it would be determinately *false* that table *T* is constructed from the relevant matter in *w*. But for this to be a

246

live option, not ruled out by the facts, it cannot be the case that it is impossible to construct two *distinct* but ordinary tables T and T' at *the very same time*, each *entirely* from *all* of a single hunk of matter. Yet this seems plainly impossible. At the very least, the prospect is every bit as paradoxical as the prospect of a pair of *mutually* possible worlds otherwise related in the same way as w_3 and w_4 in the Four Worlds Paradox. For here we would have a single world w related to *itself* in just this way. And if we consider possible tables T''' distinct from T' made from matter in the region of indeterminacy for T', presumably we should have that it is not determinately impossible that in constructing a table T' from the matter in w, we thereby construct three distinct but ordinary tables—T, T', and T''—with exactly the same matter in exactly the same place at exactly the same time. Do we not want to say that this is determinately impossible? I do.[25]

Kripke appears to include toward the view that it is of the *essence*—necessarily necessary-and-sufficient condition—of a table T that it be a table originally made from *substantially* such-and-such matter (in *substantially* such-and-such a configuration, etc.). This is a vague property. In some cases it is indeterminate whether a possible table has this property. Hence, if the property is an essence, it is a vague essence. The view that a table T has this vague essence would seem to require that for any possible world w (possible from the point of view of the actual world), and any possible table T' in w, $T' = T$ if and only if T' has this vague property in w.

Suppose we are given a world w in which some table T' is such that it is indeterminate whether T' has the property in question, i.e., suppose T' is made from matter in the region of indeterminacy of table T. It need not follow that it is vague whether $T = T'$. T' may or may not be the same table as T; we have not yet been given enough information concerning w. In fact, w represents a *class* of different worlds, with different tables T'. (Cf. the communist president case.) If we are given further that $T' = T$, then what we might infer is that the world w is not *definitely possible* from the point of view of the actual world.[26] Perhaps it is an impossible world with respect to the actual world, but perhaps it is indeterminate whether w is possible or

---

[25]But see Unger, 1980.

[26]Precisely what does follow from the biconditional depends on the details of the logic of vagueness. Insofar as the view that an entity has a vague essence leads to the consequence that identity is vague, we should again refuse the view rather than swallow its consequence.

impossible with respect to the actual world. More simply, if it is an essential property of a table *T* that it is originally constructed from *substantially* such-and-such matter, in the (weak) sense that *T* definitely has this vague property in every definitely possible world in which *T* exists, and *H* is a hunk of matter in the region of indeterminacy, then if we are given a world *w* which is not definitely impossible and in which the very table *T* is constructed from hunk *H*, then we may infer that it is indeterminate whether *w* is possible from the point of view of the actual world. This would be to acknowledge an element of vagueness. The vagueness is located not in the 'is' of identity, nor in the 'is' of constitution, but in phrases like 'substantial overlap', and hence, by way of vague essential properties, in the modal idioms themselves: 'possibly', 'could not have', etc., and the accessibility predicate of possible world discourse. If *H* is a hunk of matter differing only very slightly from the actual matter of a table *T*, then it is definitely possible for table *T* to be constructed from hunk *H*. But if *H* is a hunk of matter in the region of indeterminacy, involving an "in between" amount of overlap with the actual matter of table *T*, then it is vague whether it is possible for table *T* to have been constructed from hunk *H*. This is just to say that there is no world determinately possible (relative to the actual world) in which table *T* is constructed from hunk *H*, though there are worlds *w* in which table *T* is constructed from hunk *H*, and for which it is vague whether *w* is possible (relative to the actual world).[27]

[27]See Kripke, 1972a, p. 115, n. 57, point (3). On the account proposed here, it may be vague whether *not being constructed from hunk H* is an "essential property" of table *T*. *T* has this property in every definitely possible world in which it exists, but there are worlds which are not definitely impossible and in which table *T* exists but lacks this property.

The following definitions are hereby proposed: A proposition *p* is *definitely* or *determinately necessary* iff *p* is true in every determinately possible world and in every indeterminate world; *p* is *determinately unnecessary* iff in some determinately possible world *p* is not true; *p* is *determinately possible* iff *p* is true in some determinately possible world; *p* is *determinately impossible* iff *p* is true in no determinately possible world and in no indeterminate world; *p* is *determinately contingent* iff *p* is determinately possible but determinately unnecessary; *p* is *determinately noncontingent* iff *p* is either determinately necessary or determinately impossible. (Exercise: Prove that (*i*) if *p* is indeterminate with respect to necessity [neither determinately necessary nor determinately unnecessary], then *p* is determinately possible; (*ii*) if *p* is indeterminate with respect to possibility [neither determinately possible nor determinately impossible], then *p* is determinately unnecessary; and (*iii*) *p* is indeterminate with respect to contingency [neither determinately contingent nor

How does this account deal with Kripke's problem that the theory of vague identity was meant to solve? If *H* is an ordinary hunk of matter in the region of indeterminacy for an actual table *T*, then it is definitely possible for hunk *H* to have been made into a table. That is, there is a world *w′* which is definitely possible from the point of view of the actual world, and in which a table *T′* is constructed from hunk *H*. Is *T′* the same table as *T*? On this account, the answer must be 'no'. For we are given that *w′* is *definitely* possible with respect to the actual world, but by hypothesis there is no *definitely* possible world in which table *T* is constructed from hunk *H*.[28] On the other hand, if we are given only that there is a world *w′* which is *not definitely impossible* from the point of view of the actual world, and in which a table *T′* is constructed from hunk *H*, then there is no answer to the question 'Is *T′* the same table as *T*?' This is not

---

determinately noncontingent] iff *p* is either indeterminate with respect to necessity or indeterminate with respect to possibility.)

On the modal logic determined by these definitions, it is vague (neither true nor false) whether it is necessary that table *T*, if it exists, is not constructed from hunk *H*.

[28]This account suggests a denial of (V″). For there is a genuinely possible world *w′* in which a table *T′* distinct from *T* is the only table constructed from hunk *H* according to a certain plan. If (V″) is true, it seems to follow that it is impossible for table *T* to be the only table originally constructed from hunk *H* according to that plan, and hence that any world *w* in which table *T* is the only table originally constructed from hunk *H* according to that plan is determinately impossible. Yet on the account proposed here, there is a world *w* in which table *T* is the only table originally constructed from hunk *H* according to that plan, but which is not determinately impossible. Whether the account conflicts with (V″) depends on the logic of vagueness. What surely does follow from (V″) is that if there is a possible world (possible relative to the actual world) in which a table *x* is the only table originally constructed from hunk *H* according to the same plan, then *x* must be *T′*. The world *w* need not be taken as a counterinstance since it is not a determinately possible world. On the definitions proposed in the previous footnote, however, *w* is to be taken as a counterinstance, at least in the sense that it renders (V″) not true. Though (V″) comes out untrue on these definitions, it can still be maintained that necessarily, if a table *x* is the only table originally constructed from a certain hunk of matter according to a certain plan, then necessarily, any table that is the only table originally constructed from that hunk of matter according to that plan is the table *x* and no other:

$$\Box(x)\Box(y)\Box(P)\Box[x = (\imath z')(T(z', y) \text{ according to plan } P)$$

$$\rightarrow \Box(z)(z = (\imath z')[T(z', y) \text{ according to plan } P] \rightarrow z = x)].$$

This formulation would require only that the two worlds *w* and *w′* be mutually determinately impossible relative to each other, and allows that *w* may be indeterminate with respect to possibility relative to the actual world.

249

because identity is vague. We have not been given a single table T', but a class of worlds w' with different tables T'. In some of these worlds, though only in ones that are not definitely possible, the table constructed from hunk H is indeed the same table as T. In others, the table constructed from hunk H is not the same table as T.[29]

We have looked at three candidates for the locus of vagueness in the arguments for the Four Worlds Paradox and the intransitivity of modal accessibility—identity, constitution, and modality. We have argued that locating an element of vagueness in the last of the three appears to afford the most reasonable account. The original arguments made the assumption that there is a sharp threshold or cutoff between what matter could and what matter could not constitute a given artifact. The account just sketched in terms of regions of indeterminacy, and worlds neither determinately possible nor determinately impossible relative to one another, belies this assumption. But it does not block either of the arguments. In both cases, the general argument can easily accommodate regions of indeterminacy and vague accessibility between worlds. In the general argument for intransitivity of modal accessibility, we need only construct a sequence of worlds beginning with w_1 such that each world in the sequence is determinately possible relative to its immediate predecessor but involves some further change in the original constitution of ship a, always keeping within the limits of definitely allowable variation. Eventually we will reach a world w_n in which the original matter of ship a differs so extensively from the original matter in the initial world w_1 that w_n is determinately impossible relative to w_1. Somewhere between the two extreme worlds w_1 and w_n there may be worlds neither determinately possible nor determinately impossible relative to the initial world w_1, but that makes no difference to the argument. For we still have $w_1 \ R \ w_2 \ R \ldots R \ w_n$ but not $w_1 \ R \ w_n$. This means a failure of transitivity of R, modal accessibility, even if only

---

[29] A similar situation arises in connection with the subjunctive conditional 'If a table had been constructed from hunk H, it would (would not) have been the very table T'. (See Lewis, 1973.) On the account proposed here, for any determinately possible world w' in which a table T' is constructed from hunk H, there corresponds a nearly perfect duplicate world w which is not determinately impossible, and in which table T is constructed from hunk H, and which is, in any ordinary sense, at least as "similar" to the actual world as is w'. Whether the subjunctive conditional is true, false, or neither depends on the details of a logic of counterfactuals supplemented with a logic of vagueness.

via a region of indeterminacy.[30] Similarly, in constructing a less crude version of the Four Worlds Paradox, one can start with a pair of mutually possible worlds w_1 and w_2, each containing artifacts differing from each other in their original matter (planks) by a difference within the limits of determinately excessive (*not* allowable) variation, so that the two artifacts must be distinct. We may then work the matter of these two artifacts back toward each other, staying within the limits of determinately allowable variation, to obtain a pair of worlds w_3 and w_4 differing only in the haecceities of their respective artifacts made from the very same matter according to the very same plan.[31] But, as before, we still have no reason to

---

[30]It should be emphasized that Chandler's argument for intransitivity is no parlor trick. Any philosopher seriously concerned with the metaphysics of modality must come to terms with it. The argument appears to show that modal accessibility is intransitive, or more accurately, that on the usual and standard construal of '□', there are propositions p, e.g., that a certain table is not made originally of a certain portion of matter, such that $\ulcorner \Box p \urcorner$ is true, but something of the form $\ulcorner \Box \ldots \Box p \urcorner$ is false. I believe that it does show this. But even if the argument contains a subtle and serious flaw, there should be *something* interesting and important to be learned by exposing the error.

We have already considered, in the present section and in Section 28.2, two proposals that would block the argument for intransitivity—the inflexible essentialism of the Chisholm-Forbes counterpart theory, and Kripke's idea of vague identity—and we argued that both of these are incorrect. For two other attempts to answer Chandler's argument see Fumerton, 1978, and Odegard, 1976, p. 202. Odegard argues by replacing the inflexible essentialism of the counterpart solution with a brand of anti-essentialism concerning artifacts and their matter, whereas Fumerton seems to attack the haecceitist presupposition of the argument. It should be evident by now that I do not find either of these proposals satisfactory.

A universal $S5$ possible worlds framework has been tacitly presupposed for reasons of simplicity in a number of formulations throughout the present book, particularly in some of the modal operator discourse formalizations in Chapter Seven. However, in no instance is the main line of argument seriously affected by these considerations favoring a (vague) relational possible worlds framework.

[31]In fact, this way in which the argument can accommodate regions of indeterminacy can already be illustrated in our original example involving ship a in world w_1. We assumed that a ship of this design and structure was such that it could have been constructed from a hunk of matter differing by no more than 2% from the matter in w_1, but that a change of 3% or more must result in a different ship. It is perfectly compatible with this assumption that the interval between 2% and 3% is a region of indeterminacy. Alternatively, one can start with a pair of mutually determinately possible worlds w_1 and w_2, containing artifacts a and b, respectively, differing in their original matter by a difference just barely outside the limits of determinately allowable variation. Since the two worlds are determinately possible relative to each

suppose that w_3 and w_4 are possible relative to the initial world w_1, or to one another, and principles like (V″) might be invoked to argue that they are at least not both determinately possible relative to w_1 or to each other. (See footnote 28.) Thus it seems that the main elements of the Four Worlds Paradox, and its solution, are relatively undisturbed by the presence of vagueness and indeterminacies.

other, artifact b must be distinct from artifact a even if b is made in w_2 from matter in the region of indeterminacy for a in w_1; otherwise it would be at most indeterminate whether w_2 is possible relative to w_1. The argument then proceeds as before. This version of the argument exploits the fact that the infusion of vagueness does not, in the appropriate sense, remove the boundary lines, but introduces new ones.

The general arguments for the Four Worlds Paradox and the failure of transitivity of modal accessibility require the assumption that some alteration in the original matter of a particular artifact, even if only extremely slight (e.g., one atom), is definitely possible, while some more sweeping alteration, even if only total alteration, is definitely impossible. We need not say where the boundaries are between definitely allowable variation and indeterminacy, or between indeterminacy and definitely excessive variation, only that there are nonempty regions of determinately allowable and determinately excessive variation. In addition, the first version of the general argument of the Four Worlds Paradox depends on the further assumption that the region of indeterminacy is smaller than the region of definitely allowable variation. If the region of indeterminacy is as large as or larger than the region of definitely allowable alteration, one must resort to a sorities-type construction as in the argument for failure of transitivity. It is worth noting in this connection that in the central cases of vagueness that come readily to mind (bald vs. not bald, red vs. red-orange, embryo vs. fetus, etc.), the region of indeterminacy is relatively small when compared with the region of definiteness. In any case, the alternative version of the argument does not require the further assumption.

5

Impossible Worlds (1984)

In a recent commentary on my *Reference and Essence* (Princeton University Press and Blackwell, 1982), William R. Carter represents me as endorsing the first two of the following three modal propositions, which together constitute an inconsistent triad (following Carter's numbering scheme):[1]

(2) If a given table (ship, bicycle, etc.) x is originally constructed from a certain hunk of matter y, then it is a necessary or essential feature of x that it is originally constructed from y.

(3) If a given table (ship, bicycle, etc.) x is originally constructed from a hunk of matter (collection of material parts) y, then x could have originated from a hunk of matter z 98% (or more) of which overlaps with y; but x could not have originated from any hunk of matter z', such that less than 98% of z' overlaps with y.

(4′) If c is a material component (e.g., a molecule) of a hunk of matter y, then it is a necessary or essential feature of y that it has c as a material component.

In fact, I endorse (4′), but neither (2) nor (3). The strongest principle along the lines of (2) that I endorse is the following:[2]

(2′) If a given table (ship, bicycle, etc.) x is originally constructed from a certain hunk of matter y, then x could have been originally constructed from any hunk of matter z which is sufficiently like y (in mass, volume, composition, etc.) and which sufficiently substantially overlaps y; but x could not have been originally constructed from any hunk of matter z' which does not sufficiently substantially overlap y.

I offer (3) as one among uncountably many possible regimentations or sharpenings of (2′), one candidate for what is to count as *sufficiently substantial* overlap. I do not actually endorse (3), however, since I regard the vagueness of the phrase

---

[1] Salmon on Artifact Origins and Lost Possibilities', *The Philosophical Review*, 92, No. 2 (April 1983), pp. 223–231.

[2] Here I assume the following modal evaluation clause:

⌜$\square\phi$⌝ is true with respect to w iff ϕ is true with respect to every world determinately accessible to w and either true or neither true nor false with respect to any world neither determinately accessible nor determinately inaccessible to w.

For more on indeterminate accessibility, see *Reference and Essence*, pp. 247–252. The evaluation clause assumed here differs from the (strong) rule proposed there at p. 248, note 27.

'sufficiently substantial overlap' in ($2'$) as intrinsic to the epistemic situation. No precise principle, like (3), which removes the vagueness by substituting sharp cut-off points is knowably true. (*Cf. Reference and Essence*, pp. 240–252.)

A principle like ($2'$) is not the sort of proposition that merely happens to be true. If it is true at all, it is necessarily so. In fact, if it is true at all, then it is necessary that it is necessarily true, and it is necessary that it is necessary that it is necessarily true, and so on. From this observation, a sorites-type construction, the main idea of which has been exploited by Roderick Chisholm,[3] can be made to show that the generally accepted axiom schema of *S4* modal propositional logic,

$$\Box p \supset \Box \Box p$$

or equivalently, the notion that modal accessibility among worlds is transitive, should be rejected in its unrestricted form. Consider any hunk of matter z which is sufficiently like the hunk of matter y (in mass, volume, composition, etc.) that originally constitutes an artefact x, but which does not sufficiently overlap y. By ($2'$) it is necessary that artefact x is not originally formed from hunk z. But there is a (perhaps scattered) hunk of matter z_1 which includes some of the molecules of hunk z and which does sufficiently overlap hunk y, so that artefact x could have been formed from hunk z_1. Consider now yet another hunk of matter z_2 which includes still more of the molecules of hunk z and which sufficiently overlaps hunk z_1 (though perhaps does not sufficiently overlap hunk y). If artefact x could have been formed from hunk z_1, then (even if, in fact, x could not have been formed from z_2, still) it *might have been* that x could have been formed from z_2. Continuing in this vein, it will follow that, although it is necessary that artefact x is not formed from hunk z, still it might have been that it might have been that it might have been . . . that x is formed from z. More intuitively, if there is a possible world w_1 (possible relative to the actual world) in which artefact x is formed from hunk z_1, then there is a world w_2 possible relative to w_1 in which x is formed from z_2. Hence there is a world w_3 possible relative to w_2 in which artefact x is formed from a hunk of matter z_3 which includes still more molecules of hunk z and which sufficiently overlaps hunk z_2, and so on. Finally, there will be a world w which bears the ancestral of the accessibility relation to the actual world, and in which artefact x is formed from hunk z, though by hypothesis there is no world accessible to the actual world in which x is formed from z. World w is an *impossible world* from the point of view of the actual world.

Though the artefact x could not have been formed from hunk z, there is no reason why hunk z could not have been formed instead of hunk y into an artefact of exactly the same type and form as x in place of x itself. Thus there is a world w' possible relative to the actual world in which an artefact x', qualitatively just like x, is formed

---

[3] See his 'Identity Through Possible Worlds: Some Questions', *Noûs* 1 (March 1967), pp. 1–8; and *Person and Object* (La Salle: Open Court, 1976), pp. 148–149. Chisholm does not accept my conclusions concerning his argument, but instead rejects any principles like ($2'$) and (3) in favour of the inflexible essentialism of (2). See also Saul Kripke, *Naming and Necessity* (Harvard University Press and Basil Blackwell, 1980), at p. 51, note 18; and Hugh S. Chandler, 'Plantinga and the Contingently Possible', *Analysis* 36.1, January 1976, pp. 106–109.

from hunk z. World w' is, we may suppose, materially exactly like the impossible world w in every molecular, atomic, and sub-atomic detail. Given a complete accounting of the entire history of all of the matter in the worlds w and w', with its causal interconnections and exact configuration through time, the two worlds are absolutely indistinguishable. Atom for atom, quark for quark, they are exactly the same. Yet they must be distinct, since w' is, and w is not, a *genuinely* possible world, i.e., a world possible relative to the actual world. (*Cf. Reference and Essence*, pp. 230–40.)

Carter objects to this 'model of the situation' on the grounds of a principle of the identity of materially indiscernible worlds. If the phrase 'materially indiscernible' is understood in such a way that w and w' count as materially indiscernible, then what we have here is an example which gives the lie to this principle. It is important to notice that w and w' are indeed discernible, even if not materially discernible in this sense, and in fact discernible not only by their accessibility relations to the actual world. They also differ as regards which facts obtain in them. World w includes the fact that artefact x is formed from hunk z, whereas world w' exludes this. Some other artefact x', distinct from x, is formed from hunk z in w'. In place of Carter's principle, I propose a principle of the identity of factually indiscernible worlds, worlds in which the very same facts obtain. (*Cf. Reference and Essence*, p. 238.) I also propose a principle of the identity of *mutually accessible* materially indiscernible worlds. (*Cf. Reference and Essence*, p. 240, and p. 249, note 28.) But an unbridled principle of the identity of materially indiscernible worlds is refuted by the sorts of considerations raised here.

Is this picture of impossible worlds and mutually inaccessible materially indiscernible worlds really acceptable? There are a number of conceptions of possible worlds presently in vogue. Possible worlds are variously construed as maximal compossible sets of propositions (Robert Adams), possible total histories or states of the world (Saul Kripke, Robert Stalnaker), total ways things or the world (cosmos) might have been (David Lewis, sometimes), maximal states of affairs (Alvin Plantinga). For present purposes, these need not be regarded as competing conceptions of possible worlds (except in the case of Lewis, who usually takes nonactual possible worlds to be something like immense concrete objects, someplace far, far away). On any of these conceptions, whatever grounds there may be for believing that there really are possible worlds yield the same, or related, reasons for believing that there are impossible worlds (maximal consistent though not compossible sets of propositions, impossible total histories of the world, impossible total states of the cosmos, total ways things could not have been, etc.), for believing that there are materially indiscernible worlds (materially indiscernible total histories of the cosmos, materially indiscernible total ways for things to be, etc.), for believing the identity of factually indiscernible worlds, and the rest.

APPENDIX IV • *Modal Paradox:*
Parts and Counterparts,
Points and Counterpoints

36. MODAL PARADOX

There is a class of paradoxes that arise from the following (intuitively correct) modal principles concerning the possibility of variation in the original construction of an artifact:[1]

> If a wooden table x is the only table originally formed from a hunk (portion, quantity, bit) of matter y according to a certain plan (form, structure, design, configuration) P, then x is such that it might have been the only table formed according to the same plan P from a distinct but overlapping hunk of matter y' having exactly the same mass, volume, and chemical composition as y.

> If a wooden table x originally formed from a hunk of matter y is such that it might have been originally formed from a hunk of matter y' according to a certain plan P, then for any hunk of matter y'' having exactly the same matter in common with y that y' has, and having exactly the same mass, volume, and chemical

---

[1] I have discussed some of these paradoxes in Salmon, 1979b, pp. 722–725; Salmon, 1984a, pp. 9–11 (reprinted as Appendix III in this book, pp. 268–270); Salmon, 1984b; and in more detail in Appendix I of this book, pp. 219–252.

A version of one of the paradoxes was apparently first noted by Kripke, 1972a, p. 51 n. 18, where it is briefly discussed. Something directly akin to this paradox was also noted and discussed by Roderick Chisholm, 1973, pp. 584–586, and again in 1976, appendix B, pp. 148–149. This paradox is highly reminiscent of Chisholm's paradoxical queries concerning cross-world identity in his seminal 1967 work. I follow Graeme Forbes in calling this paradox "Chisholm's Paradox," though I am uncertain as to the propriety of the epithet. See also Quine, 1976, p. 861; and Wilson, 1959. The general solution to Chisholm's Paradox that I advocate in this volume and defend in this appendix was first proposed by Hugh Chandler (1976).

The model paradoxes have been discussed by other writers, most notably Forbes, 1983, 1984, and 1985, pp. 160–190 and passim; and Gupta, 1980, pp. 94–107. See also Carter, 1983; Forbes, 1982 and 1981a; Over, 1984, pp. 4–7; and Stalnaker, 1986.

composition as y', x is also such that it might have been originally formed from y' according to the same plan P.

(O) If a wooden table x is the only table originally formed from a hunk of matter y, then x is such that it could not have been the only table originally formed from entirely different matter, i.e., from a hunk of matter z having no matter in common with y (not even a single molecule, atom, or subatomic particle).

The last of these three modal principles, principle (O), is a nontrivial essentialist principle. It has been argued for by means of the following plausible, and perhaps more fundamental, essentialist principle concerning artifacts and their matter:

(I) If a wooden table x is such that it might have been the only table originally formed from a hunk of matter z according to a certain plan P, then there could not be a table that is distinct from x and the only table formed from hunk z according to plan P.

The argument proceeds as follows: Let x be any arbitrary wooden table that is the only table formed from its original matter y, and let z be any nonoverlapping hunk of matter. Suppose for a *reductio ad absurdum* that table x is such that it might have been the only table originally formed from hunk z instead of from hunk y. Now necessarily, every table is formed according to some plan or other. Hence there is some plan P such that table x might have been the only table formed from hunk z according to plan P. It follows directly that hunk z is such that it might have been formed into a table (some table or other) according to the very plan P, and hence z might have been so formed only once. Since table x was actually the only table originally formed from hunk y, and since hunk z might have been formed into a table only once according to

plan *P*, it might also have been that both obtained together. That is, it might have been that table *x* is the only table originally formed from hunk *y*, just as it actually was, while at the same time some table or other *x'* is the only table originally formed from hunk *z* according to plan *P*. (This is derived from a premise of the argument concerning the compossibility of certain possible states of affairs.) Of course, it is impossible for any one table to be originally formed entirely from one hunk of matter, and also originally formed entirely from some other, nonoverlapping hunk of matter. Thus, it is necessary that if table *x* is originally formed from hunk *y*, then any table formed from hunk *z* is not *x*. Hence, it is possible for there to be a table *x'* that is distinct from *x* and the only table originally formed from hunk *z* according to plan *P*. It follows by (I) and *modus tollens* that our original assumption that table *x* might have been the only table originally formed from hunk *z* is false.[2]

The first two principles cited above, taken together, imply that a certain *amount* of variation is possible in the original constitution of a table, whereas principle (0) implies that the amount of allowable variation is something short of total. A wooden table might have been originally formed from different wood, but not completely different wood; it might have been originally constructed with some different molecules, but not all. It follows that there is some threshold, some limit point—or if not a definite point, then at least some interval within which it is indeterminate—such that one more change in original constitution must by necessity result in a numerically distinct table.

It seems reasonable to suppose that the threshold consists in an interval of indeterminacy rather than a definite limit point. If a hunk of matter *y'* differs by only one molecule of wood from the original

---

[2]The argument here is derived from one given by Kripke (1972a, p. 114 n. 56; see also p. 1). The argument is analyzed in chapter 7 of this volume, pp. 196–216. For similar arguments, see Forbes, 1980 and 1981a; McGinn, 1976, p. 132. See also Sharvy, 1968 and 1983. (I should mention that I am here merely citing these works for further reference, and not endorsing the arguments or theses put forward in them.)

matter y of a table x, then clearly x is such that it might have been originally formed from y' instead of y. We have just seen an argument that if a hunk of matter z shares not even a single molecule of wood with the original matter of the table then x is such that it could not have been originally formed from z. Somewhere between these two extremes is the threshold—the minimum amount of required overlap, the maximum amount of allowable nonoverlap. The idea that this threshold amount should consist in an exact and specific number of shared molecules, or some other sort of sharp cutoff point, seems unrealistic. As with most of our concepts, our concepts of metaphysical possibility and impossibility do not seem to be quite that sharp. It seems more realistic to suppose that the threshold consists in some interval, perhaps some range of numbers of shared molecules. For any hunk of matter y' that shares a greater number of molecules with the actual matter y of the table x than any number in this range, and that is otherwise just like y, it is determinately true of x that it might have originated from y' instead of from y. For any hunk of matter z sharing fewer molecules with y than any number in the range, it is determinately true of x that it could not have originated from z. For any hunk of matter y'' whose number of shared molecules with y lies within the range, it is indeterminate—vague, neither true nor false, there is no objective fact of the matter— whether x could have originated from y'' instead of from y.

Moreover, even if there is a sharp cutoff point,[3] it seems quite unrealistic to suppose that one could ever establish—say by a philosophical proof—precisely where the cutoff point lies. Thus even if the threshold is some exact and very precise amount of overlap, from an epistemic point of view we can never be in a position to specify with adequate justification just what the threshold is—except by means of some vague locution like 'sufficiently substantial overlap'. We may assert the following:

[3]In fact, I believe that the threshold may indeed consist in a sharp cutoff point, though my approach to the modal paradoxes does not depend on this. An argument for a sharp cutoff is presented in Section 46, "The Determinacy of Identity."

(II) If a wooden table x is the only table originally formed from a hunk of matter y according to a certain plan P, and y' is any (possibly scattered) hunk of matter that sufficiently substantially overlaps y and has exactly the same mass, volume, and chemical composition as y, then x is such that it might have been the only table originally formed according to the same plan P from y' instead of from y.

(III) If a wooden table x is the only table originally formed from a hunk of matter y, and z is any hunk of matter that does not sufficiently substantially overlap y, then x is such that it could not have been the only table originally formed from z instead of from y.

It is to be understood that being exactly the same matter except for only one or two molecules counts as sufficiently substantial overlap, whereas complete nonoverlap (no shared molecules whatsoever) does not.

Paradox arises when it is noted that none of these modal principles is the sort of proposition that merely happens to be true as a matter of contingent fact. In particular, principle (II) is such that if it is true at all, it is necessarily so. Furthermore, (II) is such that if it is true at all, then it is necessary that it is necessarily true, and it is necessary that it is necessary that it is necessarily true, and so on. In fact, on the conventionally accepted system *S5* of modal propositional logic, *any* proposition is such that if it is necessarily true, then it is necessary that it is necessarily true, and it is necessary that it is necessary that it is necessarily true, and so on.

One paradox that arises from these observations I call the 'Four Worlds Paradox.' Elsewhere I have developed the paradox using the language and framework of possible-world discourse, i.e., language involving explicit reference to, and quantification over, possible worlds and possible individuals (instead of the ordinary modal locutions 'might have', 'must', or subjunctive mood). The paradox is

277

constructed by considering four distinct but related possible worlds. The Four Worlds Paradox can also be developed within modal-operator discourse, i.e., the language of the modal operators 'necessarily' or 'must', 'possibly' or 'might', and subjunctive mood. The paradox goes as follows: We consider a particular wooden table, a, with its four original legs, L_1, L_2, L_3, and L_4. Let us call the (hunk of) matter from which the table a was originally formed 'h'. The original matter of the four legs is a proper part of hunk h. Suppose for the sake of simplicity (though this is by no means essential to the argument)[4] that the threshold for table a is such that (for example) any table having the same overall plan (form, structure, design, configuration) as a is such that it might have been originally constructed using one leg different from its four actual original legs, as long as whatever other parts there are to the table (the other three original legs, the original table top, original wood screws, original glue, and so on) and the overall plan are the same. Suppose further that no table of this overall plan could have been originally constructed using two or more different legs from the actual original four. Now instead of constructing table a as he did, the artisan who constructed a might have constructed a table according to the same plan using two different table legs L_5 and L_6 in place of L_3 and L_4, keeping everything else the same—where L_5 and L_6 are qualitatively and structurally exactly like L_3 and L_4 actually are, respectively. Let us call this (scattered) hunk of matter 'h'''. Hunk h' consists of hunk h with the replacement of the matter in legs L_3 and L_4 (at the time of table a's construction) with the qualitatively identical matter in legs L_5 and L_6. By principle (III), any such (possible) table must be distinct from a itself, but there is no reason why the artisan could not have thus constructed a qualitative duplicate of a instead of a itself. In accordance with *S5* modal propositional logic, it follows by the necessitation of principle (II) that the artisan might just as well have constructed a table distinct

---

[4]See Salmon, 1979b, pp. 723–725 n. 22, and this volume, pp. 251–252 n. 31, for further details concerning the argument of the paradox.

from *a* according to the same plan using L_1, L_2, L_3, and L_6 as the four legs instead of L_1, L_2, L_5, and L_6 (keeping everything else the same), since this would involve a change of only one table leg. Let us call this hunk of matter '*h''*.' Hunk *h''* coincides exactly with hunk *h'* except for the replacement of the matter in leg L_5 with the matter in leg L_3. Now hunk *h''* also coincides exactly with hunk *h* (table *a*'s actual original matter) except for the replacement of the matter in leg L_4 with the matter in leg L_6. Since the original table a was actually formed according to the same plan from hunk *h*, it also follows by principle (II) that the artisan might have constructed *a* itself according to the same plan using the same parts—L_1, L_2, L_3, and L_6, keeping everything else the same. Thus, the artisan might have constructed *a* by shaping certain matter *h''* according to a certain plan, and he also might have constructed a table distinct from *a* by shaping exactly the same matter *h''* according to exactly the same plan. This contradicts (I).

Formally, the Four Worlds Paradox proceeds from the following set of premises, where $\ulcorner M (\alpha, \beta) \urcorner$ means $\ulcorner \alpha$ is the only table originally formed from hunk of matter β according to such-and-such a plan $\urcorner$:

$M (a, h)$	[Given]
$\Diamond(\exists x)M (x, h')$	[Given]
$M (a, h) \Box\!\sim \Diamond M(a, h')$	[from (III)]
$\Box(x) [M (x, h') \Box \Diamond M (x, h'')]$	[from $\Box$(II)]
$M (a, h) \Box\Diamond M(a, h'')$	[from (II)]
$\Diamond M (a, h'') \Box\Box(x)[M(x, h'') \Box x = a]$	[from (I)].

From these (together with the trivial assumption that necessarily, if a table is formed from some matter, then it exists, and the quantified modal logical law of the necessity of identity) the following contradiction is immediately derivable in *S5*, and even the weaker *S4*, modal logic:

(C_1) $\Diamond(\exists x)[x \neq a \And M(x, h')] \And \sim\Diamond(\exists x)[x \neq a \And M(x, h'')].$

I was once tempted by the view that this paradox is a *reductio ad absurdum* of the last premise cited above, and hence also a *reductio* of the cross-world identity principle (I). But to draw this conclusion is to miss the lesson of the paradox. Even if the last premise cited above is dropped from the list, an equally paradoxical argument can be constructed by invoking a slightly strengthened version of principle (II). To see this, let us first define the notion of a *materially complete* proposition. A proposition is materially complete if it is a complete enumeration of every particle of matter in the cosmos throughout all of a potential history of the world, as well as a complete specification of all the physical interactions and configurations of all the matter in the cosmos in exact chronological sequence throughout that potential history.

Needless to say, no materially complete proposition can be apprehended by the human mind, but of course, that is no reason to suppose that there are no such propositions. There are such propositions, and indeed one of them is true. Presumably, all true materially complete propositions are necessarily equivalent. On the modal logician's conception of propositions as sets of possible worlds (or as functions from possible worlds to truth values), exactly one materially complete "proposition" is true.

Let p be a (the) materially complete proposition that would have been true if the table a had been formed according to the same plan using leg L_6 instead of leg L_4. Notice that the materially complete proposition p surely strictly implies that some table or other is the only table originally formed from hunk h'' according to such-and-such a plan, in the sense that:

$\Box[p \ \Box(\exists x)M(x, h'')]$

Since p is a materially complete proposition that would have been true if table a had been formed from hunk h'' according to a certain

plan, it is trivial that it might have been the case both that p is true and that a is the table formed from h'' according to that plan. By an argument that proceeds exactly as before, except invoking a stronger but still intuitively correct version of (II), it also might have been the case both that p is true and that the table formed from h'' is some table distinct from a. Hence in *S4* we may derive:

$$(C_2) \quad \Diamond[p \,\&\, M(a, h'')] \,\&\, \Diamond[p \,\&\, \sim M(a, h'')].$$

This means that the question of which (possible) table is formed from hunk h'' (i.e., the question of the *haecceity* of the table formed from h'') is a question whose answer is not decided by a complete accounting of *all* the material facts in the cosmos—including the fact that hunk h'' exists as a physical unit and is table-shaped in such-and-such a particular way. This result is quite unpalatable. A table is in some obvious sense "nothing over and above" its matter and form. Perhaps some facts are underdetermined by the totality of material facts, but surely the question of whether a given actual table a is constituted by a certain hunk of matter h'' must be so determined. The fact that hunk h'' constitutes table a, if it does, is *supervenient* on a complete possible history of all the matter in the cosmos. If for some reason God had preferred to have table a originally formed from hunk h'' instead of from hunk h, once He has fixed all of the material facts—all of the facts concerning all of the matter in the cosmos—any further facts concerning which table is formed from which matter will take care of themselves. Hence, at a minimum, the following is true:

$$\Box[p \,\Box M(a, h'')] \,\vee\, \Box[p \,\Box \sim M(a, h'')].$$

This contradicts (C_2).

37. A Modal Fallacy

It is my view that both of the modal principles (II) and (III), and their multiple necessitations, are intuitively and literally true. Paradoxical conclusions are drawn from these principles by invoking defective rules of modal logic, by drawing fallacious modal inferences. Specifically, the conventionally accepted axiom of *S4* modal propositional logic,

$$\Box p \; \Box\Box\Box p,$$

or equivalently, the presumption that modal accessibility between worlds is transitive, is illegitimate and must be rejected in its unrestricted form. The modal logical system *S4* is fallacious. Its rejection invalidates a modal inference pattern critical to the Four Worlds Paradox:

$$\Box(\phi \; \Box\Diamond\psi)$$
$$\Diamond\phi$$
$$\overline{\therefore \Diamond\psi.}$$

Instead we have only the weaker inference:

$$\Box(\phi \; \Box\Diamond\psi)$$
$$\Diamond\phi$$
$$\overline{\therefore \Diamond\Diamond\psi.}$$

In particular, the hypotheses of the paradox yield the conclusion that it might have been that it might have been that a table distinct from a was originally formed from hunk h'', but they do not yield the stronger conclusion that it might have been that a table distinct from a was originally formed from h''. There is no contradiction with (I).

The primary motivation for rejecting the *S4* axiom, as applied to the

origins of artifacts (as well as other sorts of objects), is best given by means of an alternative modal paradox using a sorites-type construction, the main idea of which has been exploited by Roderick Chisholm. We begin with the same actual table a. The original matter h of table a consists of a certain number of molecules. Call this number 'n'. Now there is a finite sequence of hunks of matter, h, h_1, h_2, . . . , h_n, where each element of the sequence h_i differs from its immediate predecessor h_{i-1} only in the replacement of one molecule by a qualitatively identical but numerically distinct molecule, in such a way that the last element in the sequence, h_n, has no overlap whatsoever with h, the original matter of table a. Now by the necessitation of principle (II), each of the following necessitated conditionals is true, where $\ulcorner M(\alpha, \beta) \urcorner$ again means $\ulcorner \alpha$ is the only table originally formed from hunk of matter β according to such-and-such a plan$\urcorner$:

$$\Box[M(a, h) \Box \Diamond M(a, h_1)]$$
$$\Box[M(a, h_1) \Box \Diamond M(a, h_2)]$$
$$\bullet$$
$$\bullet$$
$$\bullet$$
$$\Box[M(a, h_{n-1}) \Box \Diamond M(a, h_n)]$$

If we head this list with the true sentence 'M (a, h)', we obtain a finite set of true premises that in $S4$ logically entail the conclusion '$\Diamond M$ (a, h_n)'. Let us call this argument (premise set plus conclusion) '(CP)', for 'Chisholm's Paradox'. The argument (CP) is $S4$-valid, and each of its premises is true. Yet by principle (III), '$\Box \sim M$ (a, h_n)' is also true. Adding this to the list of premises of (CP), we obtain a set of true premises from which a contradiction is derivable in $S4$.

One can see what is amiss with $S4$ by considering its import within the framework of possible worlds, to wit, the idea that the relation of modal accessibility between worlds is transitive. Since table a originates from hunk h in the actual world, it follows by (II)

283

that there is a world w_1 possible relative to the actual world, i.e., accessible to the actual world, in which a originates from h_1. Hence by the necessitation of (II), there is a world w_2 possible relative to w_1 in which a originates from h_2. Hence by the *double* necessitation of (II), there is a world w_3 possible relative to w_2 in which a originates from h_3, and so on. Finally, by the $(n-1)$-fold necessitation of (II), there is a world w_n possible relative to w_{n-1} in which a originates from h_n. Thus, there is a world (w_n) bearing the ancestral of the accessibility relation to the actual world and in which a originates from h_n. But by principle (III), there is no *genuinely* possible world, i.e., no world possible relative to the actual world, in which a originates from h_n. Somewhere in the sequence $h_1, h_2, \ldots, h_n$, a hunk of matter h_m ($1 < m < n$) is the first hunk to exceed the amount of allowable variation from h. Hunk h_m passes the threshold, and so, then, do all of its successors in the sequence. Hence, world w_m is not accessible to the actual world. World w_m is an *impossible world*. That is, w_m is impossible from the standpoint of the actual world, although it is possible relative to its immediate predecessor w_{m-1}, which is itself possible relative to the actual world. World w_m is a possibly possible impossible world.

Similarly, there is a world w_{2m} in which table a originates from hunk h_{2m}. World w_{2m} is possible relative to a world w_{2m-1} in which table a originates from hunk h_{2m-1}, and w_{2m-1} is possible relative to w_m, but w_{2m} is not possible relative to w_m. World w_{2m} is an impossible world that is not even a possibly possible world. It is only a possibly possibly possible world. That is, w_{2m} is a possibly possibly possible impossibly possible world.

This means that the relation of modal accessibility between worlds is not transitive. The premises of the argument (CP) are all true, but its conclusion is false. The argument (CP) is logically invalid.

If there is any defect in this illustration of the intransitivity of modal accessibility, and the consequent illegitimacy of *S4*, it is the

assumption that there is some hunk of matter h_m that is the *first* hunk in the sequence to pass the threshold. This is tantamount to the assumption that the threshold consists in some definite number of shared molecules. This assumption, however, is quite inessential to the illustration. Suppose instead that there is a range of hunks, h_k, h_{k+1}, . . . , h_{m-1}, such that for any hunk in this range, it is indeterminate—vague, neither true nor false, there is no objective fact of the matter—whether table a could have originated from it. This results in two limit points where before we had only one, and one alone is sufficient for a failure of transitivity. In the sequence of worlds w_1, w_2, . . . , w_n, each world is determinately accessible to its immediate predecessor. Furthermore, each of the worlds w_1, w_2, . . . , w_{k-1} is determinately accessible to the actual world (since it is determinately true that table a could have originated from hunk h_{k-1} or any of its predecessors), whereas each of the worlds w_m, w_{m+1}, . . . , w_n is determinately inaccessible to the actual world (since it is determinately false that table a could have originated from hunk h_m or any of its successors). Each of the remaining worlds w_k, w_{k+1}, . . . , w_{m-1} is neither determinately accessible nor determinately inaccessible to the actual work (since it is neither true nor false that a could have originated from h_k, or from h_{m-1}, or from any intervening hunk). This would mean that the accessibility relation is only *partially defined*, in the sense that its characteristic function is not total but partial. There would be a failure of transitivity *via* a region of indeterminacy, but there would still be a failure of transitivity.

Thus the modal paradoxes turn on a fallacy special to *S4* modal logic. In deriving the paradoxes in *S4*, one commits the *fallacy of possibility deletion*, inferring $\ulcorner \Diamond\phi \urcorner$ from $\ulcorner \Diamond\Diamond\phi \urcorner$, or equivalently, the *fallacy of necessity iteration*, inferring $\ulcorner \Box\Box\phi \urcorner$ from $\ulcorner \Box\phi \urcorner$. In particular, though it is necessary that table a does not originate from hunk h_m ($= h'$), it is fallacious to infer that it is necessary that it is necessary that a does not thus originate. In the Four Worlds

285

Paradox, though it might have been that it might have been that some table distinct from a is formed from hunk h'' (= h_{m-1}), it is fallacious to infer that it might have been that some table distinct from a is formed from h''.[5]

38. COUNTERPART THEORY

The primary (though not the only) rival to this approach to the modal paradoxes is derived from the modal theory of David Lewis, so-called counterpart theory. Versions of the counterpart-theoretic solu-

---

[5]If the indeterminate accessibility account sketched here is correct, principle (I) must be regarded as untrue. For suppose that z is a hunk of matter for which it is vague or indeterminate whether a particular actual table a might have been originally formed from it. Then a is not formed from z in any world determinately accessible to the actual world, though it is so formed according to some plan P in some world w_k neither determinately accessible nor determinately inaccessible to the actual world. Now it is determinately possible for a table—some table or other—to be the only table originally formed from hunk z according to plan P. Hence there is a determinately accessible world w in which some table x is formed from hunk z according to plan P. Since w is determinately accessible to the actual world, and a is not formed from z in any determinately accessible world, it follows that tables x and a must be distinct. If (I) were true, it would follow by *modus ponens* from the existence of w that in every world not determinately inaccessible to the actual world (in every world either determinately accessible or neither determinately accessible nor determinately inaccessible), no table distinct from x is the only table formed from hunk z according to plan P. (See note 14 and Section 46, below.) Yet w_k is precisely such a world in which a table distinct from x, viz., a, is the only table formed from hunk z according to plan P. Hence if there is such a world as w_k, then (I) is not true—though (depending on the details of the three-valued logic) it need not be false, since there need not be any *determinately accessible* world in which a table distinct from x is the only table formed from hunk z according to plan P. (But see note 3.)

Even if (I) is untrue for these reasons, it can be maintained that the following weakened version of (I) is necessarily true:

> (I') If a table x is the only table originally formed from a hunk of matter z according to a certain plan P, then there could not be a table that is distinct from x and the only table originally formed from hunk z according to plan P.

The necessitation of (I') is equivalent in *S4* to the necessitation of (I), though in the independent modal propositional logic *B*, the necessitation of (I') is not sufficient for the derivation of principle (0). Of course, (0) may be true nevertheless.

tion to the paradoxes have been suggested or advocated by a number of philosophers, including Hugh Chandler, Roderick Chisholm, Graeme Forbes, Anil Gupta, Saul Kripke, and Robert Stalnaker.[6]

[6]See the works cited in note 1. Differences in terminology and emphasis, as well as certain theoretical differences, tend to obscure the overall fundamental similarity among the theories advocated or suggested by these writers. In place of Lewis's terminology of 'counterparts', Chisholm employs an alleged distinction, due to Joseph Butler, between "identity in the strict and philosophic sense" and "identity in the loose and popular sense," where artifacts made in different possible worlds from different constituent molecules are, according to Chisholm, never numerically one and the very same ("in the strict and philosophic sense"), though they may be said to be "the same" in the alleged loose and popular sense. The major difference between Chisholm and the other counterpart theorists is that Chisholm does not propose to replace the standard possible-world semantic analysis of formulations in modal-operator discourse of principles like (II) by an interpretation in terms of his counterpart relation, "identity in the loose and popular sense." Thus Chisholm dissents from the formulation of (II), whereas the other counterpart theorists assent to it. Perhaps this difference is enough to disqualify Chisholm as a genuine counterpart theorist, properly so called, but I maintain that this difference is merely verbal and masks a basic agreement as to the facts (and that in this respect Chisholm is more perspicuous than the others).

Gupta occasionally uses the term 'counterpart' (p. 105) but generally prefers to speak, somewhat misleadingly, of 'transworld identity relative to a world'. On Gupta's scheme for handling the modal paradoxes, an artifact x from one world and an artifact y from another world may be said to be "identical relative to" one world w and yet not "identical relative to" another world w'. Since x is "identical" with x (itself) relative to w' (assuming x exists in w'), however, it trivially follows by Leibniz's Law, or the Indiscernibility of Identicals, that x and y are not genuinely identical—they are not one and the very same— but are two distinct artifacts. (The sort of argument just given is discussed in Section 46 below.) At most, then, x and y are merely counterparts at w, and it is at best misleading to call them *identical relative to w*. (Gupta's terminology is even more misleading than this, since he gives the title 'absolute identity' to the relation that obtains between a pair of objects when there is a world at which they are counterparts. This prompts him to make the astonishing claim that "absolute identity" is not transitive.)

I received a copy of Robert Stalnaker's "Counterparts and Identity" after the typescript of the present essay was submitted. Although I have not had the opportunity to study Stalnaker's essay carefully, several aspects of his theory seem similar to Gupta's. Stalnaker rejects my argument that his ternary, world-relative notion of "identity" is not genuine identity, but mere counterparthood, basing this rejection on the contention that there is no absolute (non-world-relative), binary notion of identity. This contention coupled with the rest of his theory, however, involves a number of serious difficulties, which I can only outline here. First, Stalnaker claims (as part of the argument that there

Forbes in particular has recently worked out many of the details of a counterpart-theoretic solution, defending it against criticisms I have made and raising objections to the intransitive-accessibility solution sketched above.[7]

Strictly speaking, one should speak of counterpart theory *with*

---

is no absolute notion of identity) that absolute truth is truth in the actual world. This gives us a notion of absolute identity: possible individuals are *absolutely identical* if and only if they are identical relative to the actual world. This notion coincides exactly with my intended notion of absolutely identity if we assume that every possible individual x is such that, actually, $x = x$. Otherwise, it is better to substitute 'some world' for 'the actual world'. One way or another, absolute identity is definable in terms of world-relative identity, and Stalnaker's theory would thus allow for my notion of absolute identity if only it admitted the notion of intraworld identity. Unfortunately, the notion that Stalnaker calls 'intraworld identity' is not identity at all. This is made clear by Stalnaker's claim (which is essential to the point of his theory) that a single individual a in the actual world (e.g., Theseus's ship) can be two distinct individuals b and c in another possible world w. He defends this claim against the charge of violating the transitivity of identity, in part, by claiming that even though in w, $b \neq c$, in the actual world, $b = c$. But in w, $c = c$. This is inconsistent with (intraworld) Leibniz's Law (which Stalnaker claims to accept), since b does not (actually) have c's (actual) property of *being identical with c in w*. If b and c (actually) differ in this respect, then whatever else they are, they are not (actually) one and the very same object—since one object cannot differ from itself in any respect. In what sense are b and c (actually) "identical," then, except in the highly misleading sense of (actually) being *distinct* counterparts of a? Given that the intraworld relation that Stalnaker calls 'identity' (actually) holds between discernible objects b and c, this relation is not, in fact, genuine intraworld identity but is merely a counterpart relation. Stalnaker purports to explain this relation as genuine identity by saying that it is the binary relation whose extension, in any possible world w, is the set of pairs $<d, d>$ such that d is in the domain of w. The phrase 'the set of pairs $<d, d>$ such that', understood in its standard set-theoretic sense, ultimately involves the notion of identity. Indeed, in fixing the extension (with respect to any possible world) of a genuine identity predicate, the notion of identity is typically invoked in order to exclude pairs of distinct objects from the extension. If Stalnaker's purported explanation thus invokes genuine identity (as it seems to), it is inconsistent (via Leibniz's Law) with his claim that the discernible objects b and c (actually) stand in the relation. If, on the other hand, the purported explanation invokes, instead, Stalnaker's world-relative notion of what he calls 'identity' relative to w, the purported explanation is highly misleading. Moreover, it is circular and does not actually fix the metaphysical intension of the relation in question.

[7]See especially Forbes, 1983, 1984, and 1985, chaps. 3, 7, appendix, and passim.

respect to a certain kind of entity, e.g., artifacts. Counterpart theory with respect to a kind *k* makes use of a binary cross-world resemblance relation, counterparthood, between possible entities of kind *k*. The counterpart relation is fixed by considerations of sufficient cross-world similarity in certain relevant respects. Since distinct possible entities of kind *k* may bear sufficient resemblance to one another across possible worlds, an individual *x* of kind *k* will have counterparts at other worlds other than itself. Typically, it is a basic tenet of the theory that each possible individual of kind *k* exists in one and only one possible world, so that a pair of counterparts existing in distinct worlds are always themselves distinct.

There are certain theoretical constraints on the counterpart relation. For example, any possible individual of kind *k* is its own counterpart at any (the) world in which it exists. Another minimal constraint is that if a possible individual *x* of kind *k* has a counterpart at world *w* that exists in *w*, then all of *x*'s counterparts at *w* exist in *w*. In the typical case, a counterpart of *x* at *w* is something that exists in *w* and (as it is in *w*) sufficiently resembles *x* as it is in its own world. Alternative versions of the theory provide for a possible individual to have a special counterpart at a world even though the counterpart does not itself exist in that world, as does Forbes's, but this happens only when the individual has no existing counterparts at the world in question. Yet another minimal constraint typically imposed is this: if a possible individual *y* is a counterpart of a possible individual *x* at a world *w*, and *y* itself has counterparts at *w* that exist in *w*, then all of *y*'s existing counterparts at *w* are also counterparts of *x* at *w*, i.e., all of a possible individual's existing counterparts at a given world are counterparts at that world of anything that the individual is itself a counterpart of at that world. This constraint can be trivially satisfied by means of the stronger constraint, typically but not always imposed, that any possible individual *y* that exists in *w* is its own sole counterpart at *w*. One condition typically not imposed, however, is transitivity. Since counterparthood is a cross-world similarity relation, and similarity is not transitive, there will be possible

individuals $x, y,$ and $z,$ such that y exists in some world w and is a counterpart of x at $w,$ and z exists in some world w' and is a counterpart of y at $w',$ but z does not sufficiently resemble x to be a counterpart of x at $w'.$

Counterpart theory (with respect to kind k) provides for a possible-world semantic theory that differs in important respects from standard Kripkean possible-world semantics for modal-operator discourse. Let us first briefly review the main ideas that differentiate standard Kripkean possible-world semantics from classical Tarskian semantics. In standard Kripkean possible-world semantics, the extensional semantic attributes—such as singular-term reference, predicate application, and sentence truth value—are relativized to possible worlds. In the case of reference and truth value, this relativization to worlds is in addition to the usual Tarskian relativization to assignments of values to individual variables. (Suppressing any reference to a model) if α is an individual variable, *the referent of* α *with respect to a world w under an assignment s,* or $Ref_{w, s} (\alpha),$ is simply the possible individual assigned to α by $s,$ i.e., $s(\alpha).$ If α is a simple individual constant, it is assigned a referent (or to use Kripke's phrase, its "reference is fixed") independently of any possible world or assignment of values to variables. Thus, simple individual constants and individual variables are *obstinately rigid designators,*[8] expressions that refer to the same thing with respect to every possible world. If Π is an n-place predicate, and $\alpha_1, \alpha_2, \ldots, \alpha_n$ are singular terms, then the atomic formula $\ulcorner \Pi (\alpha_1, \alpha_2, \ldots, \alpha_n) \urcorner$ is *true with respect to a world w under an assignment s,* or $true_{w, s},$ if and only if Π *applies with respect to w,* or *applies_w,* to the n-tuple consisting of the referents of each of the α_i with respect to w under $s,$

---

[8]A term is a *persistent* (or *persistently rigid*) *designator* if and only if it designates the same thing with respect to every possible world in which that thing exists and designates nothing with respect to all other worlds. A term is an *obstinate* (or *obstinately rigid*) *designator* if and only if it designates the same thing with respect to every possible world, whether that thing exists there or not. For more on this distinction between two types of rigid designators, see pp. 32–41 of this volume.

$<Ref_{w,s}(\alpha_1), Ref_{w,s}(\alpha_2), \ldots, Ref_{w,s}(\alpha_n)>$. The connective and quantifier cases similarly follow standard Tarskian semantics. A formula $\ulcorner\Box\phi\urcorner$ is true$_{w,s}$ if and only if ϕ is true$_{w',s}$, for every world w' accessible to w. A formula $\ulcorner\Diamond\phi\urcorner$ is true$_{w,s}$ if and only if ϕ is true$_{w',s}$, for some world w' accessible to w. A sentence is *true* (*simpliciter*) if and only if it is true$_{the\ actual\ world}$, for every assignment s.

Following the lead of Lewis, counterpart theorists typically formulate their theory in terms of translations of sentences (open or closed) involving modal operators into sentences of possible-world discourse, sentences involving explicit attribution of a counterpart relation between individuals in different worlds. This standard sort of formulation of counterpart theory may be regarded as providing a partial semantics for modal-operator discourse, in that it provides truth conditions in terms of possible worlds and counterparts for each sentence (open or closed) of modal-operator discourse. However, the semantics is only partial, since nothing is said explicitly concerning the semantics of subsentential expressions (such as singular terms and predicates) or how the truth conditions of sentences are computed from the semantics of their components. If one wishes to understand the compositional nature of the semantics of modal-operator-discourse sentences in terms of the semantics of their component expressions, one must glean this information, insofar as possible, from the translations into possible-world discourse of the modal-operator-discourse sentences in which the subsentential expressions figure. This feature of the standard formulations of counterpart theory is properly suited to a certain linguistic point of view concerning the synonymy of modal-operator-discourse sentences and the possible-world-discourse sentences giving the truth conditions of the former sentences, and the possibility of exhausting the semantics of modal-operator discourse merely by supplying possible-world-discourse sentential correlates. This point of view is disputable. Moreover, it is quite independent of the issues that separate standard possible-world theorists from counterpart theorists, and it is

quite inessential to the main philosophical ideas and intuitions that motivate counterpart theory. If a standard modal theorist adopts this point of view, he or she may easily reformulate the standard modal semantics as a set of instructions for translation of modal-operator-discourse sentences into possible-world-discourse sentences, remaining silent with respect to the compositional nature of the semantics of sentences in terms of the semantics of subsentential expressions. In order to highlight the contrast with standard modal semantics, while clearing away the unimportant differences in what has come to be the usual sort of formulations of each, it is best to reformulate counterpart theory along lines that parallel as closely as possible, within the bounds of the spirit of the philosophical motivation for counterpart theory, the usual formulation of standard possible-world semantics.

I shall do this using the notion of a *counterpart assignment*. A counterpart assignment c_w (with respect to a kind k) for a world w is a function that assigns to any possible individual i (of kind k) a counterpart of i at w, if i has any counterparts at w, and assigns nothing otherwise. If there is no counterpart of i at w existing in w, then depending on the particular counterpart theory in question, the counterpart assignment may be undefined for i, as with Lewis's theory, or it may assign the individual i to itself as its own counterpart at w, as with Forbes's. On Forbes's theory, counterpart assignments are totally defined functions.

Let us call an ordered pair of a world and a counterpart assignment for that world a *world-assignment pair*. In counterpart theory with respect to kind k, reference and truth are relativized not merely to worlds but to world-assignment pairs.[9] Thus one speaks of the referent of a singular term *with respect to* a world-assignment pair $<w$,

---

[9]A notion very similar to that of a counterpart assignment was apparently first introduced by Allen Hazen in his Ph.D. dissertation (1977). See Hazen, 1979, pp. 333–334, where analogues of counterpart assignments (there called 'representative functions') and world-assignment pairs ('stipulational worlds') are put to a use very similar to (though not exactly the same as) their use here.

c> under an assignment of values to variables *s*. Equivalently, one may speak of reference with respect to a world *w* *and* a counterpart assignment *c* for *w*, under an assignment of values to variables *s*. Similarly, one speaks of a sentence (open or closed) as being true, or as not being true, *with respect to* a world-assignment pair under an assignment of values to variables. As in standard possible-world semantics, predicate application is relativized only to worlds. The referents of simple singular terms with respect to world-assignment pairs will depend on whether the term has been assigned something of kind *k*. If α is an individual variable and *s* is an assignment of values to variables that assigns to α a possible individual not of kind *k*, then $Ref_{w,c,s}(\alpha) = s(\alpha)$. If α is an individual variable and *s* is an assignment that assigns to α a possible individual of kind *k*, then $Ref_{w,c,s}(\alpha) = c(s(\alpha))$. If α is a simple individual constant that refers to an actual individual *x* not of kind *k*, then $Ref_{w,c,s}(\alpha) = x$. If α is a simple individual constant that refers to an actual individual *x* of kind *k*, then $Ref_{w,c,s}(\alpha) = c(x)$. An atomic formula $\ulcorner\Pi(\alpha_1, \alpha_2, \ldots, \alpha_n)\urcorner$ is $true_{w,c,s}$ if and only if Π applies$_w$ to $<Ref_{w,c,s}(\alpha_1), Ref_{w,c,s}(\alpha_2), \ldots, Ref_{w,c,s}(\alpha_n)>$. A formula $\ulcorner\Box\phi\urcorner$ is $true_{w,c,s}$, if and only if ϕ is $true_{w',c',s}$, for every world *w'* and every counterpart assignment *c'* for *w'* (i.e., for every world-assignment pair $<w', c'>$). A formula $\ulcorner\Diamond\phi\urcorner$ is $true_{w,c,s}$ if and only if ϕ is $true_{w',c',s}$, for some world *w'* and some counterpart assignment *c'* for *w'* (i.e., for some world-assignment pair $<w', c'>$.)[10] Notice that the

---

[10]The truth theoretic analysis that I am formulating here by means of counterpart assignments yields some significant differences in truth value assignments to particular modal-operator-discourse sentences from Lewis's own scheme. Specifically, the following clauses for the modal operators accord better with Lewis's actual scheme:

$\ulcorner\Box\phi\urcorner$ is $true_{w,c,s}$ iff ϕ is $true_{w',c'\circ c,s}$, for every world *w'* and every counterpart assignment *c'* for *w'*;

$\ulcorner\Diamond\phi\urcorner$ is $true_{w,c,s}$ iff ϕ is $true_{w',c'\circ c,s}$, for some world *w'* and some counterpart assignment *c'* for *w'*;

where $c' \circ c$ is the composite of the assignments *c'* and *c*, i.e., the function that assigns to any possible individual x, $c'(c(x))$.

clause '*w'* is accessible to *w'* has been deleted; counterpart theory avoids the need for an accessibility relational semantics. A sentence is *true* (*simpliciter*) if and only if it is $true_{the\ actual\ world,c,s}$, for every counterpart assignment c for the actual world and every assignment of values to variables s.

The major difference between counterpart theory and standard possible-world semantics may be illustrated by means of a simple modal sentence from Chisholm' s paradox,

$$\Diamond M(a, h_1).$$

On standard possible-world semantics, this sentence is true exactly on the condition that there is a possible world (determinately) accessible to the actual world in which table a—the very table a itself—is the only table formed according to such-and-such a plan from hunk h_1 (instead of from its actual original matter h). The counterpart theorist does not admit that this condition is fulfilled. Instead, typically the counterpart theorist denies that there is any such possible world. The counterpart theorist is still able to accommodate the truth of the displayed sentence. On counterpart theory with respect to artifacts, the sentence is true exactly on the condition that in some possible world, some counterpart of a—not necessarily the very table a itself—is the only table formed according to such-and-such a plan from hunk h_1. Counterpart theory with respect to

---

Following Forbes, I am devising counterpart-theoretic possible-world semantics in such a way that '$\Diamond\Diamond F(a)$' is true exactly on the condition that a has an F counterpart at some world, so that '$\Diamond\Diamond F(a)$' is equivalent to '$\Diamond F(a)$', thus preserving *S4* modal logic. Lewis's original scheme has '$\Diamond\Diamond F(a)$' true exactly on the condition that a has a counterpart that itself has an F counterpart at some world. Since coutnerparthood is not transitive, this condition may be fulfilled though a itself has no F counterpart at any world. Lewis's scheme thus fails to preserve *S4*, since '$\Diamond\Diamond F(a)$' is weaker than '$\Diamond F(a)$'. This separates Lewis motivationally from theorists such as Forbes, who invoke counterpart theory precisely to retain *S4* modal logic in the face of the modal paradoxes. This does not mean, however, that Lewis himself blocks the Four Worlds Paradox in the same way as the accessibility solution. See note 20.

artifacts thus assigns a different truth condition to the sentence, one whose fulfillment seems beyond doubt.

In effect, counterpart theory replaces the intransitive accessibility relation with an intransitive counterpart relation. There are glaring technical differences between the two types of solutions to the modal paradoxes, however. (There are glaring motivational differences as well. The motivation for counterpart theory, as a solution to the modal paradoxes, is discussed in Section 40 below.) First, certain intuitively correct premises involved in Chisholm's Paradox are counted unequivocally true on the accessibility solution but cannot be thus accommodated on counterpart theory (as I have formulated it). Consider the argument (CP). Suppose again that in the sequence of hunks of matter, h_1, h_2, ..., h_n some one hunk h_m is the first in the sequence to pass the threshold. Then on counterpart theory with respect to artifacts, the premise

$$(P_m) \quad \Box[M(a, h_{m-1}) \Box \Diamond M(a, h_m)]$$

will not be true, since there is a world w_{m-1} in which a counterpart of table a is formed from hunk h_{m-1}, whereas at any world in which a table is formed from hunk h_m, that table, though a counterpart of the counterpart of a at w_{m-1}, is not a counterpart of a itself. (A similar situation obtains if the threshold is vague and there is a range of hunks h_k, h_{k+1}, ..., h_{m-1} for which it is indeterminate whether a possible table formed from one of these hunks is a counterpart of a.) Thus whereas the accessibility solution blocks (CP) by counting it logically invalid, counterpart theory with respect to artifacts (as I have formulated it) blocks (CP) by counting it logically valid but unsound.[11]

Another glaring difference between the two solutions to the modal paradoxes is brought out in their respective treatments of the Four Worlds Paradox. Although counterpart theory with respect to artifacts

---

[11]On Lewis's original scheme, (P_m) comes out true. See note 10 above.

is able to accommodate *S5* modal propositional logic, in so doing it foregoes certain valid inferences of standard quantified *S5* modal logic. In particular, it is able to accommodate the truth of the necessitation of the modal principle (II), and of certain sorts of instances of it, like the fourth premise of the Four Worlds Paradox,

$$\Box(x)[M(x, h') \Box \Diamond M(x, h'')].$$

In standard quantified modal logic, it follows from this together with the result

$$\Diamond(\exists x)[x \neq a \ \& \ M(x, h')]$$

and the trivial truism

$$\Box(x)\Box[M(x, h'') \Box (\exists y)(y = x)]$$

that

$$\Diamond\Diamond(\exists x)[x \neq a \ \& \ M(x, h'')].$$

Counterpart theory with respect to artifacts invalidates this inference and thereby blocks the paradox. The accessibility solution, on the other hand, allows the inference, but invalidates further inference by possibility deletion. Thus both solutions to the Four Worlds Paradox count the argument of the paradox invalid, though on distinctly different grounds. Similarly, counterpart theory with respect to artifacts accommodates

$$\Box(x)[M(x, h_{m-1}) \Box \Diamond M(x, h_m)]$$

while blocking the inference from this together with '$\Box[M(a, h_{m-1}) \Box (\exists x)(x = a)]$' to the (CP) premise ($P_m$) displayed above.

In the general case, if counterpart-theoretic possible-world semantics is devised in such a way as to preserve *S5* modal propositional logic together with the philosophical institutions that motivate the theory, it foregoes the following modal version of universal instantiation, valid in standard quantified modal logic:

(MUI) $\Box(x)\phi_x$

$\therefore \Box[(\exists x)(x = \alpha) \Box\phi_\alpha]$,

where α is a simple individual constant or individual variable other than 'x', ϕ_α is just like ϕ_x except for having free occurrences of α wherever ϕ_x has free occurrences of 'x', and ϕ_x may contain occurrences of modal operators. This deviation from standard quantified modal logic prevents the derivation of paradoxical conclusions from the necessitation of (II).[12]

---

[12]Specifically, counterpart-theoretic possible-world semantics, as I have devised it, invalidates the inference

$$\frac{\Box(x)[G(x) \supset \Diamond F(x)]}{\therefore \Box(Exists(a) \supset [G(a) \supset \Diamond F(a)])},$$

since it may be that every possible individual that is G in its own world has an F counterpart at some world, and that a has an existing G counterpart at some world, though a itself (as opposed to its G counterpart) has no F counterpart at any world. The trouble with this instance of (MUI) arises from the nesting of modalities in the conclusion. Lewis's original scheme validates this instance of (MUI), but as noted in note 10 above, it does not preserve *S4* modal logic. In a sense, then, the counterpart theorist is faced with a choice between *S4* and such instances of (MUI). Forbes chooses the former, Lewis the latter. Standard quantified *S5* modal logic validates both *S4* and (MUI). (Neither version of counterpart theory validates all instances of (MUI).)

Of course, counterpart-theoretic possible-world semantics can be artificially made to capture as much standard quantified *S5* modal logic as desired by placing further constraints on the counterpart relation. Standard modal logic emerges as the special case of counterpart-theoretic modal logic where counterparthood is identity. It is the philosophical motivation for counterpart theory, and the consequent explication of counterparthood in terms of sufficient cross-world similarity in certain respects, that requires the nontransitive and one-many nature of counterparthood.

39. MODAL PARADOX AND SORITES

Each of the necessitated conditional premises of the argument (CP) is equivalent in *S4* to an unnecessitated material conditional, so that the argument may be recast in *S4* into the standard form of a sorites argument in classical propositional logic:

$$(CP)' \quad \Diamond M(a, h)$$

$$\Diamond M(a, h) \ \Box \Diamond M(a, h_1)$$
$$\Diamond M(a, h_1) \ \Box \Diamond M(a, h_2)$$
$$\bullet$$
$$\bullet$$
$$\bullet$$
$$\Diamond M(a, h_{n-1}) \ \Box \Diamond M(a, h_n)$$

$$\therefore \ \Diamond M(a, h_n).$$

Forbes emphasizes this feature of Chisholm's Paradox and argues that the paradox should be treated in a manner exactly parallel, or as closely as possible, to a contemporary treatment of the standard propositional sorites paradox, such as the paradox of the short person:

Anyone only 5 ft. tall is short.
If anyone 5 ft. tall is short, then so is anyone 5

ft. $\dfrac{1}{1,000,000}$ in. tall.

$\bullet$

$\bullet$

$\bullet$

If anyone 5 ft. 11 and $\dfrac{999,999}{1,000,000}$ in. tall

is short, then so is anyone 6 ft. tall.

∴ Anyone 6 ft. tall is short.

Standard sorites paradoxes arise from vagueness in some key expression or concept. In the case of the paradox of the short person, the key term is the adjective 'short', which is clearly, true of anyone (or at least, any adult human) only five feet tall, clearly false of anyone six feet tall, but neither clearly true nor clearly false with respect to a range of heights in between. Now one extremely plausible way of diagnosing the problem with this sorites argument is as follows. Assuming that the first premise of the argument is true and that the conclusion is false (its negation true), somewhere down the list of the 12 million conditional premises to the argument—in fact, at least twice, and most plausibly, a large number of times down the list—a conditional premise is neither true nor false. For somewhere down the list there is a conditional with a true antecedent but a consequent neither true nor false, followed by a sequence of conditionals with both antecedent and consequent neither true nor false, followed finally by a conditional with an antecedent neither true nor false and a false consequent. Each of these premises is itself neither true nor false. Thus the classical sorites argument in propositional logic is formally valid but unsound. Not all of its premises are true, even if none are strictly false.[13]

---

[13]Here and throughout this essay I am ignoring the possibility that semantic terms like 'true' and 'false' might themselves be vague or have partially defined semantic characteristic functions. If 'true' and 'false' are themselves vague or otherwise partially defined, a simple atomic sentence may suffer from second-order vagueness or second-order failure of truth value, in that the sentence may be, say, determinately untrue though it is indeterminate (vague, neither meta-true nor meta-false, there is no objective fact of the matter) whether the sentence is false or not. Similarly, a sentence that is determinately not false may be neither determinately true nor determinately untrue. The possibility of higher-order vagueness does not directly affect the main points I wish to make concerning the modal paradoxes and sorites paradoxes, and further discussion of this phenomenon in the present essay would introduce unnecessary complications. Notice that it is still reasonable

Appendix IV

The solution to the modal paradoxes offered in Section 37 above allows for a treatment of (CP)' exactly parallel to this. In particular, the critically vague term involved in (CP)', if any (see note 3), is the accessibility predicate of possible-world discourse, and thereby the possibility operator '◇' occurring throughout (CP)'. A sentence $\ulcorner◇\phi\urcorner$ is true (*simpliciter*) if and only if ϕ is true with respect to some world determinately accessible to the actual world. The same sentence is false (*simpliciter*) if and only if ϕ is false with respect to every world determinately accessible to the actual world and untrue—either false or neither true nor false—with respect to every world neither determinately accessible nor determinately inaccessible to the actual world.[14] The intransitive accessibility account allows that there may be a hunk of matter h_k such that table a originates from it in some world neither determinately accessible nor determinately inaccessible to the actual world, but does not originate from it in any determinately accessible world. If this is so, $\ulcorner◇M(a, h_k)\urcorner$ is neither true nor false. Hence at least two of the conditional premises of (CP)' will be neither true nor false, just as in the paradox of the short person. Insofar as it is desirable for a solution to (CP)' to parallel as closely as possible a contemporary solution to the classical propositional sorites paradox, the indeterminate accessibility solution does exactly what is desired.

More important than this, the accessibility solution severs the

---

to count a classical propositional sorites argument unsound, even if the sequence of sentences making up the antecedents and consequents of the conditional premises (e.g., sentences of the form $\ulcorner$Anyone of height h is short$\urcorner$) run the full gamut from determinately true to determinately not false but neither determinately true nor determinately untrue, to determinately neither true nor false, to determinately untrue but neither determinately false nor determinately not false, and finally to determinately false. Some of the conditional premises would have to be counted determinately not false while neither determinately true nor determinately untrue, but still others should be counted determinately not true. For it would be most reasonable to count a conditional determinately neither true nor false, and hence determinately untrue, whenever its antecedent is determinately not false but its consequent is determinately neither true nor false, and similarly whenever its antecedent is determinately neither true nor false and its consequent is determinately untrue.

[14]Here I assume the three-valued modal semantics I put forward in Salmon, 1984b, p. 114 n. 2, rather than that found in this volume, p. 248 n. 27.

alleged equivalence between (CP) and (CP)', and in fact, the original modal argument (CP) comes out differently in a very important respect from a standard sorites argument. Unlike the premise set of the propositional sorites argument (CP)', all of the premises of the original argument (CP) are determinately true, whereas its conclusion is determinately false. This reflects a crucially important difference between Chisholm's Paradox and the standard sorites paradox. It is important to remember that Chisholm's Paradox, as well as the Four Worlds Paradox and others belonging to the same class, are paradoxes of *modality*. Chisholm's Paradox is not a paradox in classical propositional logic, but a paradox in modal logic. The key feature of Chisholm's Paradox—the feature of it that makes it a peculiarly modal paradox—is its essential use of *nested modalities*. It proceeds from the observation that the truth of the modal principle (II) is no accident but is a necessary truth, thus yielding the nesting of modal operators in the modal premises of (CP). The intransitive-accessibility solution to Chisholm's Paradox properly distinguishes between the original argument (CP) and the propositional recasting (CP)', the latter being a familiarly valid but unsound argument in classical propositional logic and the former an interestingly invalid argument in modal logic. It is a critical defect in the counterpart-theoretic solution (as well as other rivals to the intransitive-accessibility solution) that it is blind to the crucial differences that separate the two cases. The modal paradoxes, as they naturally arise in pondering essentialist doctrines of the sort put forward in principle (III) (and as they did in fact arise in Chisholm's pioneering queries on the subject), are peculiarly modal in that they involve nested modality and depend upon the fallacy of possibility deletion, or equivalently, the presumption that accessibility between worlds is transitive. The counterpart-theoretic solution, in attempting to reduce the modal paradoxes to "the previous case" of standard sorites paradoxes such as the paradox of the short person, recommits the same fallacy and, in so doing, fails to recognize the rightful status, and consequently the proper lesson, of the modal paradoxes.

40. Some Shortcomings of Counterpart Theory

The fundamental defect of the counterpart-theoretic solution to the modal paradoxes is revealed when considering the motivation for invoking counterpart theory in attempting to solve the paradoxes. If Chisholm's Paradox is to be regarded on the model of the paradox of the short person, one must ask what term or expression involved in the former plays the role of the crucially vague term 'short' involved in the latter.

It cannot be expression 'a' itself. In fact, it is not in the least bit clear what it would mean to say that a proper name—or an individual constant such as 'a', which functions as a proper name—is "vague," unless it means that 'a' is ambiguous or nonreferring. We may pretend, for present purposes, that the name 'a' unambiguously refers to a particular table. The paradoxes still arise. It is even less clear what it would mean to say that the table a itself is vague, unless it means that the table has a vague *boundary*, in the sense that with respect to certain molecules at the periphery of the table, it is vague—indeterminate, neither true nor false, there is no objective fact of the matter—whether they are or are not constituents of the table. But vagueness in the table's boundary is not at issue here; the modal paradoxes would arise even if tables came with sharp boundaries.

Nor is there any relevant vagueness in the term 'table', or in the property of being a table. No doubt there are things such that it is vague whether they are to count as tables (as opposed to, say, counters or chests), but we may take it that a itself is a clear and central case of a table. The paradox still arises.

Nor is there any relevant vagueness in the hunks of matter h, h_1, . . . , h_n. We may suppose that these are precisely given, with an exact accounting of every molecule included and the exact configuration of their totality. The paradox still arises. Nor is there any relevant vagueness in the relational concept of a table x being originally formed from a hunk of matter y according to such-and-such a plan. Wherein, then, does the vagueness reside?

One might try looking at the matter thus: In the short person paradox,

there is a sequence of heights, 5 ft., 5 ft. $\dfrac{1}{1,000,000}$ in., . . . , 6 ft., such

that, though each height is precise and exact enough in itself, for some of these precisely delineated heights it is vague whether someone of that exact height counts as being *short*. Similarly, in the case of Chisholm's Paradox, we have a sequence of hunks of matter, $h_1, h_2, \ldots, h_n$, each precisely given, and a corresponding sequence of worlds, $w_1, w_2, \ldots, w_n$, such that in any world w_i there is a table a_i just like a except that it is originally formed from hunk h_i instead of from a's original matter, h. This sequence of possible tables, a_1, $a_2, \ldots, a_n$, plays the role analogous to that of the sequence of heights in the short person paradox. Each is precisely given, though for some it is vague whether the table still counts as being a or not. In the actual world, there is also a table just like a originally formed from hunk h. This table is a itself. In world w_n, the table a_n formed from hunk h_n is definitely not a, since by principle (III) there is no genuinely possible world in which a is formed from h_n. With respect to certain worlds w_k intermediate in the sequence between the actual world and w_n, it is vague—indeterminate, neither true norfalse, there is no objective fact of the matter—whether the table a_k formed from hunk h_k in that world is or is not the very table a from the actual world. To use the contemporary vernacular, what is indeterminate is whether a_k has a's haecceity—the property of being identical with a—in w_k. Thus the vague concept involved in Chisholm's Paradox would appear to be that of *being identical with a in a possible world*, or more simply, *possibly being a*. More specifically, since the name 'a' is itself nonvague, the relevant vague concept involved would appear to be the relational concept of cross-world identity, or that of possible identity, expressed by '$\Diamond x = y$'. Evidently, this vagueness

traces to vagueness in the very concept of identity itself. The ultimate source of the vagueness involved in Chisholm's Paradox thus appears to be the 'is' of identity.

Kripke, apparently having reasoned along lines similar to these, concludes that a counterpart-theoretic approach may be useful in dealing with the vagueness of identity in Chisholm's Paradox. He says that

> perhaps, . . . given certain counterfactual vicissitudes in the history of the molecule of a table, *T*, one may ask whether *T* would exist, in that situation, or whether a certain bunch of molecules, which in that situation would constitute a table, constitute the very same table *T*. . . . In concrete cases we may be able to answer whether a certain bunch of molecules would still constitute *T*, though in some cases the answer may be indeterminate. (1972, pp. 50–51)

In a footnote to this passage, he writes:

> There is some vagueness here. If a chip, or molecule, of a given table had been replaced by another one, we would be content to say that we have the same table. But if too many chips were different, we would seem to have a different one. . . . Where the identity relation is vague, it may seem intransitive; a chain of apparent identities may yield an apparent nonidentity. Some sort of 'counterpart' notion . . . may have some utility here. One could say that strict identity applies only to the particulars (the molecules), and the counterpart relation to the particulars 'composed' of them, the tables. The counterpart relation can then be declared to be vague and intransitive. . . . Logicians have not developed a logic of vagueness. (p. 51 n. 18)

There are a number of difficulties with this motivation for the counterpart-theoretic approach. Kripke's idea seems to be that where

(the characteristic function of) the concept of identity is undefined, it may facilitate a semantic investigation if the identity concept is represented in the metalanguage by means of a surrogate relation, counterparthood, which is vague and intransitive. Now it may indeed facilitate a semantic investigation into the logic of a vague term or predicate such as 'bald' to consider various regimented or sharpened surrogates or approximations to the vague concept, precisely defined—say, in terms of an exact number of strands of hair on the top portion of the head per square inch of surface area. One might thus verify the validity of the inference $\ulcorner \alpha$ has a full head of hair $\therefore$ α is not bald$\urcorner$. But Kripke is proposing that an allegedly vague concept, identity, be investigated in terms of *another vague concept*, counterparthood. It is difficult to see how there is anything to be gained in representing one vague concept by means of another. If our problem is that we lack a logic of vagueness, we can no more treat the latter than we can the former. If our purpose is to investigate the logic of identity among tables, surely we are better off sticking with genuine identity and doing the best we can, than turning our attention elsewhere only to find the same obstacles arise there.

Perhaps Kripke committed a slip of the pen here and meant to declare the counterpart relation to be *non*vague and intransitive—as opposed to genuine identity among tables, which it represents and which (we are to suppose) is vague but transitive. For example, one might define a relation of counterparthood in such a way that any possible table is a counterpart of itself, i.e., of a determinate self, whereas for any pair of possible tables a_i and a_j for which it is either false or vague (neither true nor false) that they are identical, neither counts as a counterpart of the other. This counterpart relation would thus play the facilitating role of a sharpened or regimented approximation to identity among tables and other artifacts.

Even when Kripke's proposal is modified in this way, it seems confused. It is quite unclear what it means to say that strict identity does not "apply" to tables. Suppose there is a possible table a_k such

that a and a_k are neither determinately identical nor determinately distinct. Then on this interpretation of Kripke's proposal, a is a counterpart of a (itself) but not of a_k. It follows directly by Leibniz's Law, or the Indiscernibility of Identicals, that a and a_k are distinct, contradicting the hypothesis. (A similar argument applies if counterparthood is defined so that a and a_k are counterparts.)

The defender of this proposal may protest that within the counterpart-theoretic framework, one is barred from saying anything about the identity or distinctness of a and a_k. One can speak only about the cross-world similarity relations between a and a_k; one must settle for the weak claim that a and a_k are not counterparts. But the Leibniz's Law inference cries out to be drawn; if a is a counterpart of a but not of a_k, then a has a counterpart that a_k does not have. Whether we are allowed to say so or not, it follows that a and a_k *cannot be* one and the very same object and *must be* distinct. Our refraining or being prohibited from saying so does not make it any less true.

When the truth is spoken, incoherence is the result. Consider again the sequence of possible tables $a, a_1, a_2, \ldots, a_n$. Kripke's remarks concerning this sort of situation are highly compressed, and his exact intent is unclear. He says: "Where the identity relation is vague, it may *seem* intransitive; a chain of *apparent* identities may yield an *apparent* nonidentity" (emphasis added). Presumably, if "the identity relation is vague," then things that are *apparently* identical (or *apparently* distinct) need not be *determinately* identical (or *determinately* distinct). A pair of objects x and y may *appear* to be identical (or distinct) when in reality, there is no objective fact of the matter as to their identity (or their distinctness). Perhaps Kripke's view, then, is this: (i) any table in the sequence $a, a_1, a_2, \ldots, a_n$ appears to be identical to its immediate successor in the sequence; (ii) the initial table a and the final table a_n appear to be distinct; but (iii) in reality, for any pair of tables a_i and a_j where $i \neq j$, there is no objective fact of the matter concerning their identity or distinctness.

In that case, Kripke's view would involve rejection of both the

modal principles (II) (since a and a_1 only *appear* identical) and (III) (since a and a_n only *appear* distinct). This is not a very satisfactory solution to the modal paradoxes. Both (II) and (III) are intuitively correct, even if it is vague what is to count as "sufficiently substantial overlap." In fact, if Kripke's view is that it is vague—or indeterminate, or neither true nor false, or there is no objective fact of the matter—whether tables a and a_n are distinct, then his view involves rejection of the modal principle (0), a principle that is both weaker than (III) and precisely formulated in a way that (III) is not. It would be difficult, if not impossible, to reconcile this consequence of Kripke's view with his attempt in the very same work to provide "something like proof" for principle (0), or a principle directly like it. (See note 2 above.)

Another possible view might be that in the sequence of possible tables a, a_1, a_2, . . . , a_n, each element is determinately identical with its immediate successor, though there is some range of elements, a_k, a_{k+1}, . . . , a_{m-1}, that are each neither determinately identical with nor determinately distinct from the initial element a, whereas the next element in the sequence, a_m, and all of its successors are determinately distinct from the initial element a. However, this is equally incoherent. If a and a_{k-1} are determinately identical, then they are *one and the very same*, and if a_{k-1} and a_k are determinately identical, then they are also one and the very same. But then there is only one table here. Which table? Well a, aka a_k. Tables a and a_k are one and the very same after all; they are determinately identical. Conversely, if a_m is determinately distinct from a, yet determinately one and the very same table as a_{m-1}, then a_{m-1} must be determinately distinct from a after all. Moreover, if each element in the sequence and its immediate successor are one and the very same, then what we have is simply an n-ary sequence of table a taken n times in a row. It is quite literally impossible for some element in this sequence to be distinct from a. Conversely, if any table in the sequence fails to be determinately identical with the initial table a, then the sequence is

not simply the *n*-ary sequence of *a* taken *n* times in a row. Hence it is impossible for each element in the sequence to be one and the very same as its immediate successor.

The idea that the identity relation is vague, in the sense that its characteristic function is undefined for certain pairs of concrete objects like tables, is itself incoherent. In fact, it is provable that the identity concept, or the 'is' of identity, is totally defined for every pair of individuals. The proof, which was foreshadowed in the arguments just given, goes as follows: Suppose, on the contrary, that there is a pair of individuals, *x* and *y*, for which the 'is' of identity is undefined—a pair to which neither the predicate 'are one and the very same' nor its negation 'are not one and the very same' correctly applies. Then this pair $<x, y>$ is quite definitely *not* the same pair as the reflexive pair $<x, x>$, since the 'is' of identity—or the predicate 'are one and the very same'—does correctly apply to the latter. That is, the pair $<x, x>$ is an element of the extension of the 'is' of identity (the class of ordered pairs of which the predicate is determinately true), whereas the pair $<x, y>$ is not; hence, they are distinct. It follows by standard ZF set theory that $x \neq y$. But then, contrary to the hypothesis, the 'is' of identity is defined for the pair $<x, y>$. The 'is' of identity is determinately false of the pair; its negation correctly applies. The general form of this argument can be applied to a variety of philosophical issues concerning identity.[15]

In fact, this brief argument also proves that the concepts of identity within a possible world, i.e., intra-world identity, and of cross-world identity (and by analogy, identity at a time and identity over time) are also totally defined. For each is definable in terms of absolute, unrelativized identity as follows:

$x =_w y =_{def.} x = y$.

x in w_1 is identical with *y* in $w_2 =_{def.} x$ exists in

w_1 & *y* exists in w_2 & $x =_{w_1} (\imath z)[y =_{w_2} z]$.

---

[15]The proof just presented that identity is nonvague is discussed further in Section 46.

Perhaps most important, Chisholm's Paradox and the other modal paradoxes do not even involve the concept or relation of identity. The paradoxes *can* be formulated in terms of possible identity or cross-world identity, but they can just as easily be formulated without identity. In fact, the 'is' of identity does not occur in either (CP) or (CP)'—not once, not anywhere. If (CP) and (CP)' constitute a paradox of vagueness, the vagueness must reside in one or more of the terms actually used in the formulation. Since the identity predicate does not even occur, if there is any vagueness, it must reside elsewhere. It is a mistake to see Chisholm's Paradox as stemming from vagueness in identity.

Forbes's motivation for his counterpart-theoretic approach to Chisholm's Paradox is somewhat different from Kripke's, though he seems to mislocate the vagueness in the same place. He writes:

> [There] is no sharp distinction between those sums [of matter] which could, and those which could not, constitute [the table *a*]. Given that there is no fuzziness in the boundaries of particular sums of wood or in the constitution relation, it seems that this vagueness must arise from an underlying vagueness in the concept of possibly being identical to [*a*]; however, in standard [possible-world] semantics, such vagueness could only be represented by vagueness in [*a*'s cross-world] identity conditions, and a solution of the paradox in which we think of identity as vague would be rather unappealing. But [it] does make sense to think of *similarity* as being vague, in the sense of admitting *degrees*. . . . [The] counterpart relation is fixed by similarity considerations—in the present context, similarity of design and constituting matter. (1984, p. 174)

Forbes's overall argument appears to be this: The orginal argument (CP) is equivalent in *S5* to (CP)', a standard propositional sorites-type argument; hence it is simply a special case of a general

and familiar sort of paradox of vagueness. Since the vagueness cru-
cially involved in (CP)' does not reside in the hunks of matter h_1, h_2,
..., h_n or in the relation of being a table formed from such-and-such
matter, it must reside in the concept of possibly being *a*. On the stan-
dard possible-world semantic analysis of modal-operator discourse,
this would mean that there is vagueness in the identity relation itself.
But the idea that identify is vague is "rather unappealing" as a solu-
tion to Chisholm's Paradox. Counterpart theory provides an alterna-
tive possible-world semantic analysis of modal-operator discourse in
which the vagueness of possibly being *a* is derived not from vague-
ness in identity, but from vagueness in a relation of similarity, the
relation of counterparthood. Therefore, a counterpart-theoretic
approach should afford a superior solution to Chisholm's Paradox.

This motivation for the counterpart-theoretic solution, though
apparently different from Kripke's, is defective in a related way. As
I have already noted, neither the argument (CP) nor its alleged equiv-
alent (CP)' involves the concept of identity, and hence neither
involves the concept of possibly being identical with *a*. If (CP)' con-
stitutes a paradox of vagueness, the vagueness must reside else-
where, in some concept essentially involved in the argument.

In fact, despite Forbes's motivational remarks, in his formal treat-
ment the vagueness is indeed located elsewhere. Specifically, by
invoking a counterpart theory in which the counterpart relation is
vague, Forbes formally locates the vagueness involved in (CP) in a
certain second-order modal concept: the concept of a property's
being such that *a* might have had it. Formally, the crucially vague
expression involved in (CP)', according to Forbes's formal treat-
ment, is $\ulcorner \Diamond \ldots a \ldots \urcorner$, or $\ulcorner$it might have been that $a \ldots\urcorner$; the crucially
vague concept is that designated by '$\lambda F \square F(a)$'.

Forbes's formal treatment may be correct in imputing vagueness
to this modal locution, for if there is any vagueness relevantly
involved in Chisholm's Paradox, it can be only in such locutions as
this. However, it is not at all true that standard possible-world

310

semantics can accommodate the vagueness of this locution only by treating identity as vague. In fact, even if identity is (incoherently) regarded as vague, that would not be sufficient to impute vagueness to the locution in question, since this locution does not involve the identity predicate. It involves only the sentential possibility operator and the proper name (individual constant) '*a*'. We have already seen that the name '*a*' is not a source of vagueness. Hence, if there is any vagueness relevantly involved in the modal paradoxes, it resides in the modal operators themselves, and the modal operators are precisely where Forbes's formal treatment ultimately locates the vagueness upon which the paradoxes turn.

We have also already seen that nothing so radical as a departure from standard possible-world semantics in favor of a counterpart-theoretic semantics is called for in order to accommodate vagueness in the modal operators. Standard possible-world semantics can accommodate the relevant vagueness in the modal operators in precisely the way I have suggested: one should treat the accessibility relation between worlds as itself vague (its characteristic function partially defined), so that certain pairs of worlds are neither determinately mutually accessible nor determinately mutually inaccessible. When fully worked out, this involves intransitivity in the accessibility relation via a region of indeterminacy, and hence an abandonment of *S4* modal logic in favor of something weaker or independent (such as the modal system *B*). This approach affords a solution to the modal paradoxes that accommodates vagueness precisely where it must arise, if anywhere, and it does so within the framework of standard possible-world semantics without resorting to the entirely unnecessary and unjustified tack of invoking counterparts in place of cross-world identities. This approach also recognizes a crucial difference between the modal paradoxes and the standard paradoxes of vagueness: the former turn on a fallacy special to modal logic—the fallacy of possibility deletion, or equivalently, the fallacy of necessity iteration.

The counterpart-theoretic approach is not merely unnecessary and unjustified. It is positively misleading and logically distinctly counterintuitive. I shall develop these criticisms each in turn.

41. More Shortcomings of Counterpart Theory

Counterpart theory appears to provide an alternative to standard possible-world semantics that is able to accommodate modal principles like (0), (I), (II), and (III), and their multiple necessitations, within an *S5* framework (i.e., maintaining an equivalence accessibility relation) without generating the paradoxes. Yet as it is typically intended, counterpart theory with respect to artifacts accommodates precisely the opposite of (II): if a wooden table x is originally formed from a hunk of matter y, and y' is any hunk of matter distinct from y, then even if y' substantially overlaps y and is otherwise just like y, x is such that it *could not* have been originally formed from y' instead of from y. The reason for this is that, as it is typically intended, counterpart theory with respect to artifacts includes the basic tenet that possible artifacts formed in their respective possible worlds from distinct (even if substantially overlapping) hunks of matter are always themselves distinct (though they may be mutual counterparts). Thus if x is a wooden table originally formed from a hunk of matter y, and y' is a hunk of matter even only one atom or molecule different from y, the counterpart theorist with respect to artifacts would typically deny that there is a genuinely metaphysically possible scenario, a genuinely possible world, in which the one and only very table x—that very table and no other—is formed from y' instead of from y. The counterpart theorist will insist that, strictly speaking, if we are ever to have one and the very same table x—that very table and no other—existing in a counterfactual scenario that might have obtained, x must be originally formed in that scenario from exactly the same matter, atom for atom, quark for quark, right down to the

tiniest of subatomic material components. For this reason, counterpart theory with respect to artifacts is, at bottom, a particularly inflexible brand of essentialism. The counterpart theorist with respect to artifacts can mouth the words 'x might have been formed from y' instead of from y', thereby seeming to advocate (II). But in counting this remark true and therefore assertible, the counterpart theorist means to be committed to nothing more than the availability of a possible scenario in which some table or other sufficiently similar to x—not necessarily x itself—is formed from y'. The counterpart theorist thus says one thing and means another.[16]

Forbes has responded to this objection by claiming that

> whether or not [counterpart] theory admits contingency (of the table x's original matter] . . . turns only on whether or not it [counterpart theory] is consistent with the truth of [the sentence 'x is formed from and might have existed without being formed from y'], and by this criterion, counterpart theory admits contingency beyond all question. (1984, p. 179)

This response involves a confusion—or perhaps an equivocation—between two distinct senses in which a theory may be said to "admit" or accommodate a principle or proposition.[17] A theory accommodates a proposition p in the primary sense if the theory embraces p itself, that is, if the proposition p is included as a part of the theory (or at least as a logical consequence of the theory in combination with uncontroversial premises). A theory may be said to accommodate a proposition p in a secondary sense if the theory (or the theory in combination with uncontroversial premises) logically entails the metatheoretic proposition that some particular sentence ϕ

---

[16]See pp. 232–238.

[17]I am concerned here with theories in the ordinary sense of the word. A theory in this sense is not merely a set of expressions closed under a special syntactic relation, but something more along the lines of a set of fully *interpreted* sentences, or a set of propositions, closed under genuine logical consequence.

is true, where φ is in fact a formulation of, or expresses, the proposition *p*. These two kinds of accommodation should be sharply distinguished. Counterpart theory with respect to artifacts can indeed accommodate modal principles like (0), (I), (II), and (III) in the secondary sense. But this sort of accommodation is deceptive, since as it is typically intended, counterpart theory with respect to artifacts fails to accommodate the critical principle (II) in the primary sense. Consider, by analogy, the following simple theory: (1) A table's exact original matter is always an essential feature of the table; (2) snow is white; and (3) the sentence 'Any particular wooden table might have been formed from metal instead of wood' means in English that snow is white. Call this theory '*T*'. (To dispel the appearance of inconsistency, imagine the theory *T* being formulated in Chinese.) The theory *T* can hardly be said to admit contingency of original matter in any relevant sense, though it does accommodate (II) in the secondary sense. Like counterpart theory with respect to artifacts, the theory *T* avoids the modal paradoxes by *rejecting* the modal principle (II)—not the *formulation* of (II) given above, but the proposition (II) itself. It may not be entirely futile, but it would be a difficult matter indeed to argue the merits of the doctrine of contingency of original matter with a proponent of *T*. The advocate of *T* will apparently join in singing the praises of (II), but the agreement is merely verbal.

Forbes argues that to see counterpart theory on this model, as an inflexible essentialist theory that misrepresents the meanings of modal-operator-discourse formulations of principles such as (II), is

> to think of the extensional sentences of [possible-world discourse] as having some meaning given independently [of modal-operator discourse]. . . . But this conception of [the relation between the two types of discourse] is not very plausible. . . . The threat is that . . . we would have to . . . identify possible worlds with logical constructions of actual entities; and . . . [this identification] has recently been

shown [by Alan McMichael] to be problematic. It seems better to think of the meanings of [sentences of possible-world discourse] as being given by those of the modal [-operator-discourse sentences] themselves (so far as this is possible). . . . [In giving the meanings of sentences of possible-world discourse by means of sentences of modal-operator discourse] it would be up to the theorist himself to decide just how to proceed, given his purposes. . . . But from this starting point, one cannot think of the sentences of either [standard or counterpart-theoretic possible-world] semantics as yielding perspicuous representations of the 'real' meanings of the modal [-operator-discourse] sentences. . . . Yet Salmon's criticism makes sense only if we think of [possible-world discourse] in these unlikely ways. (ibid., pp. 179–180)

Forbes's conception of the nature and content of possible-world semantics raises large issues concerning the enterprise of semantics generally, issues too broad in scope to be debated adequately in the present forum. It is worth noting, though, that Forbes's conception of the nature of possible-world semantics is distinctly implausible when extended to temporal semantics for tensed discourse, though Forbes has also suggested that some sort of temporal-counterpart theory may be useful in solving temporal paradoxes analogous to the modal paradoxes.[18] Semantics for tensed discourse usually employs the notion of a time *t*—perhaps a moment of time or an interval of time—and the relation of earlier-later between times. A semantics for tensed discourse can also be developed using the idea of an *instantaneous total state of the cosmos*, or what I shall call an *i.s.*, and the relation of temporal precedence between successive instantaneous states (assuming no instantaneous state of the cosmos is ever repeated). Using instantaneous states of the cosmos in place of times better emphasizes the analogy between temporal semantics and pos-

---

[18]Forbes, 1983, p. 252; see also p. 258 n. 27.

sible-world semantics. In *i.s.* semantics for tensed discourse, the semantic attributes of reference, application (of a predicate), and truth value are relativized to *i.s.*'s. On the natural semantic development, a sentence of the form 'It has been the case that ϕ' is true with respect to an *i.s. i* if and only if ϕ is true with respect to some *i.s.*, or some succession of consecutive *i.s.*'s, that precede *i*. Similar clauses may be given for other temporal operators ('it is going to be the case that', 'it has always been the case that', and so on). Now perhaps the meaning of the phrase 'instantaneous total state of the cosmos' is such that it can be explained, or is in fact learned, only by means of tense or other temporal operators; perhaps not. In either case, the phrase has a relatively clear meaning, and contrary to the spirit of Forbes's remarks, this meaning determines the correct correspondence between a sentence of temporal-operator discourse and the expression of its truth condition in *i.s.*-discourse, not vice versa. It is not the prerogative of the semanticist to devise whatever semantic clauses suit his or her philosophical interests and temperament.

Consider, for example, the sentence 'Bill has been baptized'. On the natural semantic development, this sentence is true in English with respect to the present *i.s.* if and only if Bill is baptized in some prior succession of consecutive *i.s.*'s. It is quite incredible to suppose that a philosopher particularly fond of the idea of cross-time resemblance is free to select some other truth condition for this sentence more to his or her liking. A temporal-counterpart theorist might tell us that on his or her theory, the tensed sentence 'Bill has been baptized' is translated into the following sentence of *i.s.*-discourse:

> In some succession of consecutive instantaneous total states
> of the cosmos that precede the present instantaneous total
> state, someone bearing such-and-such a resemblance to Bill,
> as he presently is, is baptized.

The claim that this sentence means simply that Bill has been baptized is bizarre. Even if Bill is now remarkably like his great-grand-

father used to be, the fact that his great-grandfather was baptized has no bearing semantically on the truth in English of 'Bill has been baptized'. Of course, one could decide to use the *i.s.*-discourse sentence displayed above in such a way that it is, in effect, a semantically unstructured *idiom*, one that means simply that Bill has been baptized (in the way that the phrase 'kick the bucket' means to die), but such a decision involves a misleading and radical departure from English. The point of introducing such misleading idioms into semantics would be utterly mysterious. Why not use the original straightforward formulations in temporal-operator discourse?

The fact is that sentences of *i.s.*-discourse do not function in *i.s.* semantics as unstructured idioms, whether standard *i.s.* semantics or temporal-counterpart-theoretic *i.s.* semantics. On the contrary, it is the very internal semantic structure of *i.s.*-discourse sentences that makes *i.s.*-discourse suitable for the enterprise of doing a systematic semantics for a tensed language. In fact, the very existence of the theory of instantaneous states and cross-time counterparthood offered by the temporal-counterpart theorist gives the lie to the claim that the meaning of an *i.s.*-discourse sentence (such as the one displayed above) is fixed by its alleged analogue in tensed discourse ('Bill has been baptized'). Rather, the meaning of an *i.s.*-discourse sentence is fixed in the usual way, by the meanings of its components—including the meanings of 'instantaneous total state of the cosmos', 'precede', 'present', and 'resemblance', as they arise in formulating the temporal-counterpart theory. Thus the *i.s.*-discourse sentence displayed above has a meaning that involves the temporal-counterpart theorist's concept of cross-time resemblance. It cannot mean the same thing as the tensed discourse sentence 'Bill has been baptized', for the proposition that Bill has been baptized involves no concept of resemblance, and hence it does not involve the particular resemblance concept given in the temporal-counterpart theory and expressed in the proposed *i.s.*-discourse translation. The same is true if reference to persisting objects is replaced with reference to tem-

317

poral stages of persisting objects, and if cross-time resemblance is replaced with a notion of spatiotemporal continuity.

Consider now a contemporary follower of Heraclitus who holds that one cannot step into the same river in the same spot twice—i.e., in two different instantaneous states of the cosmos—because new water is continuously flowing through. The contemporary Heraclite (perhaps unlike Heraclitus himself) believes that, in general, the matter that constitutes an object (e.g., the water in a river) is a permanent and unchanging feature of the object. A contemporary Heraclite may devise an elaborate temporal-counterpart theory with respect to material objects to make it possible to "speak with the vulgar"—to utter sentences like "The Mississippi River once had cleaner water flowing through it than it now has"—but then this clever Heraclite does not mean by this sentence what the rest of us mean, or what the sentence itself means. Any such pronouncement in tensed discourse by this philosopher is merely a verbal camouflage. When the Heraclite says 'The Mississippi is the same river today as yesterday', he or she does not mean the word 'same' in the "strict and philosophic sense," but rather in what he or she believes is a "loose and popular sense," i.e., as a word for temporal counterparthood.[19]

The phrase 'possible world' may not be as clear in meaning as the phrase 'instantaneous total state of the cosmos', but there are a number of conceptions of possible worlds presently in vogue, each of which is clear enough to substantiate my labeling of counterpart theory as a particularly inflexible brand of essentialism. Possible worlds are variously construed as maximal compossible sets of propositions (Robert Adams), total histories the world might have had (Saul Kripke), maximal states of affairs that might have obtained (Alvin Plantinga), total states the cosmos might have been in (Saul Kripke, Robert Stalnaker), total scenarios that might have obtained (myself). For present purposes, these need not be regarded as competing conceptions of possible worlds. If the phrase 'possible world'

---

[19]Cf. Chisholm's doctrine of *ens successiva* in Chisholm, 1976, pp. 89–113, 145–158.

is unclear in meaning, any of these clearer phrases may be substituted.[20] It is of course true that each of these explications, of what a possible world is involves notions from modal-operator discourse: *possible, compossible,* or *might have.* The notion of a possible world is defined in terms of concepts like *might have,* rather than vice versa. In this sense, the meanings of sentences of possible-world discourse are not "given independently" of modal-operator discourse. But they do have meaning, and just as in the case of tensed and *i.s.*-discourse, the meanings of sentences in possible-world discourse determine the semantic clauses for modal-operator discourse, and not the other way around. It is not the prerogative of the semanticist to stipulate whatever semantic clauses suit his or her philosophical interests and temperament. The sentence 'Bill might have been a robot' is true if and only if there is a possible scenario—or a possible history, or a possible state of affairs, or a possible state of the cosmos—in which Bill is a robot. The availability of a possible scenario in which not Bill but something rather like Bill in such-and-such respects is a robot is entirely irrelevant.

As in the case of tensed and *i.s.*-discourse, the very existence of the counterpart theorist's theory of possible worlds and counterparthood as a relation of cross-world similarity gives the lie to Forbes's claim about what fixes the meanings of the possible-world-discourse sentences that allegedly give the truth conditions of sentences in modal-operator discourse. The meanings of possible-world-discourse sentences are fixed in the usual way, by the meanings of their grammatical components—including the word 'counterpart', as it arises in the

---

[20]One might also employ a conception, due to David Lewis, of possible worlds as *ways things might have been.* My reason for not including this in the list is that Lewis himself (usually) takes a way-things-might-have-been-but-are-not to be something like an immense concrete object someplace "far, far away," in another dimension of the total cosmos, rather than a way the cosmos might have been, i.e., a possible state of the cosmos. If Lewis insists on this conception of a possible world, strictly speaking his version of counterpart theory is not a brand of essentialism at all, nor is it even relevant to modality in general. It is a fantastic cosmological theory.

counterpart theorist's formulation of his or her theory. Another indication of this is the fact that Forbes relies on possible-world discourse, rather than on untutored modal-operator-discourse intuition, as the court of last arbitration to determine the fine detail of which inferences in modal-operator discourse are to count as valid and which are to count as invalid. The very enterprise of a systematic possible-world semantics for modal logic would be impossible if the sentence giving the truth-in-a-model condition for a particular object language sentence has its meaning fixed by the object language sentence itself.

The fact that counterpart-theoretic semantics misinterprets modal-operator discourse is made evident by the logic the former imposes on the latter. We have already seen that if counterpart theory is devised in such a way as to preserve *S5* modal propositional logic, it typically invalidates certain special cases of an intuitively valid modal variant of universal instantiation, (MUI), which permits the inference from 'Necessarily, everything is ϕ' to 'Necessarily, if α exists, then it is ϕ', where α is a simple singular term. (See note 12.) The misinterpretation of modal-operator discourse is made even more plain if a predicate for the intra-world analogue of counterparthood is added to the latter. For if the counterpart theory includes the usual constraint that all of a possible individual's existing counterparts within a given world are counterparts at that world of anything that the object is itself a counterpart of at that world, then the theory validates the intuitively fallacious inference from

$$\Diamond(\exists x) \, [x \text{ is a counterpart of } a \,\&\, F(x)]$$

to

$$\Diamond F(a).$$

The validity of this inference in counterpart-theoretic modal logic illustrates the weak interpretation placed on simple possibility sen-

tences such as 'Bill might have been a robot'. Normally, if someone were to utter this sentence, he or she would mean something considerably stronger than, if not entirely independent of, whatever may be entailed by the claim that there might have been a robot counterpart of Bill.[21]

42. THE SOLUTION REFINED

The various explications of possible worlds given in the preceding section support the legitimacy of the idea of an impossible world, as well as the intransitive-accessibility account of the modal paradoxes. Just as there are such things as maximal compossible sets of propositions, there are also such things as maximal consistent but not compossible sets of propositions. If there are such things as total histories the world might have had, maximal states of affairs that might have obtained, and total states the cosmos might have been in, then there are also such things as total histories the world could not have had, maximal states of affairs that could not have obtained, and total states the cosmos could not have been in. Some of these impossible worlds are such that they might have been possible worlds instead of impossible worlds; their modal status as possible or impossible is a contingent feature of them. In fact, among the impossible worlds are

---

[21]Cf. this volume, pp. 234–235. Forbes has responded to this objection by claiming, in effect, that the alleged validity of the inference in question does not violate any relevant logical intuition. See Forbes, 1984, p. 182, and 1985, p. 180. This might be taken as an indication that Forbes does not mean what the rest of us mean by 'Bill might have been a robot'. Most of us understand this sentence in such a way that it is true if and only if there is a scenario that might have obtained—or a history the world might have had, or a state of affairs that might have obtained, or a state the cosmos might have been in—in which Bill himself is a robot. Contrary to Forbes, we have a strong logical intuition that the proposition that Bill himself might have been a robot, whether true or false, is no logical consequence of any proposition to the effect that there might have been a robot counterpart of Bill—unless counterparthood is just identity.

those that might have been possible, those that could not have been possible but might have been such that they might have been possible, those that could not have been such that they might have been such that they might have been possible but might have been such that they might have been, and so on, perhaps to infinity. In any case, for some fairly large finite number n, there are worlds that are not possible in the nth degree (not possibly possibly . . . ($n - 1$ times) . . . possible), but that might have been, i.e., they are possible in the (n + 1)th degree. Consider, for example, the possible total scenario (history of the world, and so on) w_1 in which everything is just as it actually is except that the table a is formed from the hunk of matter h_1 instead of from hunk h (and whatever other differences are required by this difference in order to ensure genuine possibility). The total scenario (history, and so on) w_2, that is just like w_1 except that table a is formed from hunk h_2 instead of from h_1, is a possible scenario relative to scenario w_1. That is, in scenario w_1, scenario w_2 is a possible scenario. Eventually, there is a total scenario w_m that is not possible relative to the actual total scenario, i.e., that is not a genuinely possible scenario, but that might have been. That is, w_m is possible in the second degree, but not in the first. Similarly, as we have seen, the total scenario w_{2m}, in which table a is formed from hunk h_{2m}, is possible in the third degree, but not in the second, and hence not in the first. Even the total scenario w_n, in which table a is formed from entirely different matter, is possible in some sufficiently large degree, though presumably it is not possible in only the second or third degree.

Thus far I have ignored the fact that certain sentences may be neither true nor false, perhaps in virtue of a false presupposition, as with the occurrence of a nonreferring definite description (e.g., Russell's 'The present king of France is bald'), or in virtue of the occurrence of a vague predicate (e.g., 'Louis is bald', where Louis has enough hair on his head so that he is not determinately bald but not enough hair so that he is determinately not bald). When we take note of this

fact, it emerges that possible worlds are not *maximal* or *total* in the ordinary sense. For example, the proposition that the present king of France is bald is arguably neither true nor false in the actual world, so that the set of true propositions includes neither this proposition nor its negation (the proposition that the present king of France is not bald). If the actual world is just the set of true propositions, then a possible world may be a compossible set of propositions that falls short of being genuinely maximal. Similarly, if the actual world is the true history of the world, or the total state the cosmos is in, and so on, then since the true history of the world, and the total state the cosmos is in, include nothing that determines that the present king of France is bald and also nothing that determines that the present king of France is not bald, a possible world may fall short of being *total* in the sense of deciding 'yes' or 'no' on every possible question of fact. Still, of course, a possible world must approach maximality or totality as closely as possible. A possible world must be *maximal* or *total* in the weaker sense that for any proposition or question of fact p left undecided, there must be enough propositions or questions of fact decided (e. g., that there is no present king of France, or that the number of hairs on the top portion of Louis's head per square inch of surface area is n) to determine that there is no objective fact of the matter concerning p.

This observation supports the feasibility of the indeterminate-accessibility account of the modal paradoxes sketched in Section 37 above. A total (in the weak sense) scenario w' is accessible to a total scenario w if and only if it is a fact in w that w' might have obtained. If the notion of possibility is itself vague, there will be total (in the weak sense) scenarios w_k such that the actual total scenario includes nothing about whether w_k might have obtained or not. World w_k would thus be neither determinately accessible nor determinately inaccessible to the actual world, in the same way that some people are neither determinately bald nor determinately not bald in the actual world.

Forbes objects to these conceptions of what possible worlds are by endorsing a criticism, due to Alan McMichael,[22] that a theory of such entities as maximal compossible sets of propositions or maximal states of affairs that confines itself to things that actually exist—an actualist theory of such entities—is problematic. McMichael's criticism, very briefly, is this. The following sentence involving nested modalities is true:

> *S*: It might have been the case that there exists someone who: (*a*) does not actually exist; (*b*) is bald; and (*c*) might have existed without being bald.

Following the standard approach rather than the counterpart-theoretic approach, *S* is true if and only if there is a possible world *w* in which there exists an individual *x* such that (*a*) *x* does not exist in the actual world; (*b*) *x* is bald in *w*; and (*c*) there is a world *w'* accessible to *w* in which *x* exists but is not bald. McMichael argues that this truth condition apparently cannot be fulfilled within an actualist theory of possible worlds. Suppose for example that possible worlds are identified with maximal compossible sets of states of affairs. Then in order for *S*'s truth condition to be fulfilled, it seems there would have to be one such set *w* that includes the state of affairs of there existing an individual *x* who does not actually exist and who is bald, and another such set *w'* that includes the states of affairs of *x*'s existing and *x*'s not being bald. But since *x* does not actually exist, there are no such states of affairs as *x*'s existing or *x*'s not being bald, and hence no such set as *w'*.

The argument here is fallacious, though exposing the fallacy is a delicate matter. No such set as *w'* is required to exist for the truth of *S*. Exactly what is required is the existence of a maximal compossible set *w* of states of affairs that includes the complex state of affairs of there existing an individual *x* such that: (*a*) the state of

[22]McMichael, 1983.

affairs of x's existing does not actually obtain; (b) x is bald; and (c) there is a maximal compossible set of states of affairs w' that includes the states of affairs of x's existing and x's not being bald. This in turn requires the existence, and the possibly obtaining, of the state of affairs of there existing some individual or other who is bald, whose existence does not actually obtain, and whose existence while not being bald might have obtained. But it in no way requires the existence of either the state of affairs of this nonactual individual's existence or of his or her not being bald.

We may put the matter this way: Suppose that possible worlds are maximal compossible sets of propositions. Now it has been observed by a number of philosophers, including McMichael, that within an actualist framework, a set of possibly true propositions may be maximal (in either the strong or weak sense) and yet may include some particular existential generalization without including any singular instance of it. This occurs when the existential generalization is such that no actual entity yields an instance that is possibly true. For example the proposition expressed by '($\exists x$) [x does not actually exist]', though false, is such that it might have been true. Since there is no actual entity that can serve as the relevant constituent of a possibly true singular instance of this existential generalization, however, there is no singular instance that is possibly true. Now in order for S to be true, there must be a maximal (in the weak sense) compossible set w of propositions that includes the proposition expressed by the sentence:

($\exists x$) [x does not actually exist & x is bald & ($\exists w'$) (w' is possible & the proposition that x exists and is not bald $\in w'$)].

As was just indicated, w will include no singular instance of this proposition, since there are none to be included. More importantly, however, the sentence displayed above is equivalent to the following:

($\exists w'$) [w' is possible & ($\exists x$) (x does not actually exist & x is bald & the proposition that x exists and is not bald ϵ w')].

This sentence also expresses precisely the sort of existential proposition that is possibly true but has no possibly true singular instance. What the truth of S requires is the existence, and the possible truth, of this existential proposition; it does not require the existence of any singular instance of it.

43. VAGUENESS AND MODAL PARADOX

My criticisms of counterpart theory are independent of the logic of vagueness that may be supplied to supplement the theory. In fact, the logic of vagueness is all but irrelevant to the main idea behind a counterpart-theoretic approach to the modal paradoxes. Forbes proposes treating the counterpart relation as itself vague and a matter of degree. Essentially the same account results from speaking of *determinate counterparts* in place of counterparts *simpliciter*—where, if it is indeterminate to some degree whether x is a counterpart of y, then it is determinately true that x is not a determinate counterpart of y.

Following J. A. Goguen,[23] Forbes proposes to treat the sort of vagueness found in concepts like that of being short or that of being similar by means of infinitely many *degrees of truth and falsehood* in place of the conventional all-or-nothing dichotomy of truth and falsehood. Accordingly, on Forbes's theory, a sentence containing a vague term may be wholly true, almost wholly true, more true than false, equally as true as false, more false than true, almost wholly false, or wholly false. Degrees of truth and falsehood are represented by means of the real numbers between 0 and 1, inclusive, where 1 represents complete truth, 0 represents complete falsehood, and the sum of the degree of truth of a sentence and its degree of falsehood (the degree of truth of its negation) is 1.

[23]Goguen, 1969.

Many find the idea of a sentence being (unambiguously) partly true and partly false grating. Truth and falsehood appear to be mutually exclusive absolutes; nothing "partly false" is genuinely and literally true in the ordinary sense. But it would be a mistake to conclude that the concept of degrees of truth and falsehood is utterly without merit in the logic of vagueness. To illustrate: suppose there are two men, Smith and Jones, for whom it is vague—indeterminate, neither true nor false, there is no objective fact of the matter— whether either is bald. Ordinarily, though neither of the two men has little enough hair to qualify as determinately bald, one of the two, say Smith, will be "balder" than the other, in the sense that Smith has proportionately less hair on the top portion of his head per square inch of surface area than does Jones. Neither is determinately bald, but Smith is "closer" to being determinately bald than Jones is. Although the adjective 'bald' is neither true nor false of both men, it is closer to being true of Smith than it is to being true of Jones. The sentence 'Smith is bald' is closer to being true than is the sentence 'Jones is bald', though neither sentence is true (and neither is false). One may decide to put this another way by saying that both sentences partake of a certain "degree of truth" less than the "maximal degree," and that the first is "more true" than the second. It does not follow, of course, that the first sentence is true *simpliciter*—any more than Smith's being taller than Jones entails that Smith is tall.

Similarly, though the proportion of hair on Smith's head does not fall squarely into either the category *bald* or the category *not bald*, in all likelihood it is closer to one end of the scale than to the other. Suppose that Smith is such that if he were to lose just a very few more strands of hair, he would become determinately bald rather than indeterminate with respect to baldness, whereas if he grew as many strands of hair, he would remain indeterminate with respect to baldness. Then the sentence 'Smith is bald', though neither true nor false, is closer than its negation to being true; it is closer to being true than it is to being false. One might put this by saying that it is "more

true than false," though strictly speaking, of course, it is neither. This interpretation provides significance to the notion of "degrees of truth" in the logic of vagueness.

The important point about this construal of a degrees-of-truth approach should not be obscured by the somewhat misleading jargon of a sentence being "more true than false" or "more false than true." A sentence that is true *simpliciter* is now being said to be "wholly" or "completely" true, or true "to the maximum degree," and a sentence that is false *simpliciter* is now being said to be "wholly" or "completely" false, or false "to the maximum degree." A sentence said to be only "partly true," or "less than but almost wholly true," is not true at all, and a sentence said to be "less than but almost wholly false" is not false at all. On the construal I am suggesting of the degrees-of-truth approach, the classical three-way division among *true*, *false*, and *neither true nor false* is built into that approach—as maximal truth, maximal falsehood, and everything in between. The range of degrees of truth between maximal falsehood and maximal truth, exclusive, are nothing more than gradations of the traditional category of *neither true nor false*, so that classical three-valued logics emerge as subtheories of analogous degrees-of-truth approaches. If a sentence is said to be "more true than false," or "almost but not quite wholly true," it is neither true nor false, though in the sense sketched above it is closer to being true than to being false.

This interpretation of the degrees-of-truth machinery evidently clashes with Forbes's intent. First, Forbes has denounced the traditional three-way division among *true*, *false*, and *neither* as arbitrary, whereas on the construal suggested this division is embedded in the degrees-of-truth approach.[24] More important, Forbes's definition of

[24]Forbes, 1984, p. 177. On the other hand, Forbes is willing to identify his notions of maximal truth and maximal falsehood with the traditional truth values of classical two-valued logic (1985, p. 170).

The possibility of vagueness of infinite order (see footnote 13 above) suggests an alternative interpretation of the degrees-of-truth semantics for vagueness. On my suggested construal of the degrees-of-truth approach, second-order vagueness can be accommo-

validity in the logic of vagueness does not accord well with the suggested construal of the nature of the truth value status represented by real numbers between 0 and 1. Forbes calls an argument or inference pattern 'valid', roughly, if in any model, its conclusion is at least as true as the least true of its premises (more accurately, if in any model, the degree of truth of the conclusion is at least as great as the greatest lower bound of the degrees of truth of the premises). This leads him to brand *modus ponens* an invalid inference pattern, since in his logic of vagueness, a conditional that is neither (wholly) true nor (wholly) false may be closer to being true (have a "greater degree of truth") than either its antecedent or its consequent taken individually. Forbes calls the inference pattern of *modus ponens* 'the fallacy of detachment' and blames the standard sorites paradoxes on this alleged fallacy. He sees the choice between the accessibility solution to the modal paradoxes and the counterpart-theoretic approach as a choice between rejecting *S5* modal logic while consequently treating the two arguments (CP) and (CP)' differently, on the one hand, and rejecting *modus ponens* while treating the two arguments equivalently, on the other. Since *modus ponens* must be rejected in any case, Forbes argues, the counterpart-theoretic approach is superior to the accessibility approach. It retains *S5* modal logic while allegedly reducing (CP) to a familiar paradox of vagueness in classical propositional logic.

Can it be that *modus ponens* is a fallacious inference pattern and that this is the fallacy involved in the traditional sorites paradoxes, such as the paradox of the short person? I can think of no inference pattern whose validity is more obvious than *modus ponens*. Rather than place myself in the hopeless position of Achilles, though, I will

---

dated by allowing that a sentence may be determinately greater than 0 (or determinately less than 1) in truth value status while it is indeterminate whether the sentence takes on the value 1 (or 0) rather than some real between 0 and 1. As with first-order vagueness, the degrees-of-truth approach with indeterminacy allows for finer distinctions than the simply three-valued approach with indeterminacy. Still, on this construal, the latter approach is completely embedded within the former.

say here only that the validity of *modus ponens* is certainly more intuitively obvious and compelling than the alleged validity of the *S4* axiom of modal logic, or equivalently, the inference pattern of necessity iteration (possibility deletion). If the choice were as Forbes sees it, the accessibility approach should be the winner beyond all question!

In fact, though, Forbes has posed a false dichotomy. An inference pattern is properly *valid* if and only if it is truth-preserving, i.e., if and only if for every instance, its conclusion is true in every model in which its premises are true. This is the proper notion of validity even in the logic of vagueness. In a degrees-of-truth logic of vagueness, as I have proposed construing it, an inference pattern is valid (properly so called) if and only if it preserves "complete truth" or "truth to the maximum degree." By this criterion, *modus ponens* is unquestionably valid even on the degrees-of-truth account. Why place the blame for paradoxes of vagueness on *modus ponens*? In fact, the traditional sorites argument in classical propositional logic is perfectly valid. What goes wrong in a standard sorites paradox, such as the paradox of the short person, is not that the argument is invalid, but that it is unsound. Not all of the conditional premises are (wholly) true. At least two are neither true nor false. In the terminology of the degrees-of-truth approach, at least two conditional premises are "partly true and partly false," or "less than wholly true but true to some degree." The paradox of the short person is dissolved by noting that one should not attempt to establish conclusions by reasoning from premises that are untrue—even if they may be said to be "almost wholly true" in the sense sketched above. Almost is simply not good enough.

The intransitive-accessibility account rejects *S4* and accommodates both of the modal principles (II) and (III) in both the primary and secondary senses, whereas the counterpart-theoretic approach, as I have devised it, retains *S5* but fails to accommodate (II) in the primary sense. This is the real choice. The paraphernalia of degrees-of-truth, the alleged loose and popular sense of 'identity', cross-

world counterparts, so-called identity from the point of view of a particular world, and the rest, tend to obscure the point.

44. TWIN WORLDS

On the intransitive-accessibility account that I advocate, there are distinct yet purely qualitatively identical worlds in which the very same matter exists in exactly the same configuration, and in which all matter undergoes exactly the same physical processes, down to the finest detail, throughout all of time. This is not quite the same as the apparent conclusion (C_2) of the second version of the Four Worlds Paradox, which is surely unacceptable. It is open for the accessibility theorist to argue (as I have elsewhere) that any two distinct such worlds are mutually inaccessible and are not both (determinately) accessible to the actual world. In modal-operator discourse, the accessibility account yields the following conclusion, where p is a (the) materially complete proposition that would have been true if table a had been formed from hunk h'' [$= h_{m-1}$] rather than from hunk h:

$$(C_3) \qquad \Diamond[p \ \& \ M(a, h'')] \ \& \ \Diamond\Diamond[p \ \& \ {\sim}M(a, h'')].$$

Some philosophers have objected to this conclusion on the basis of a principle of the identity of materially indiscernible worlds, i.e., worlds in which the same materially complete proposition is true.[25]

---

[25]See, for example, Carter, 1983, pp. 228–229; Forbes, 1984, pp. 182–184, and 1985, p. 165n.

Forbes in particular has objected that the acceptance of this conclusion is incompatible with a general metaphysical principle concerning identity facts and the concept of identity—a principle that entails an extreme, cross-world version of the Identity of Indiscernibles (from which the identity of indiscernible worlds is derivable), and that, according to Forbes, provides the ultimate justification for the essentialist principle (III). This is the reductionist or supervenience principle that all facts about the numerical identity or distinctness of a pair of objects, x and y—including facts of cross-time and cross-

Conclusion (C_3) is in fact perfectly compatible with any reasonable version of the Identity of Indiscernibles. Moreover it can be modified to show that the principle of the identity of materially indiscernible worlds in fact contradicts classical Indiscernibility of Identicals. This can be seen through consideration of another conclusion correctly obtainable from the assumptions of the Four Worlds Paradox:

$$\Diamond\Diamond M(a, h').$$

If table a had been formed from hunk h'' instead of from hunk h, then it *would have been* possible for it to have been formed from hunk h' instead of from hunk h''. Let p' be a (the) materially complete proposition that would have been such that it would have been true if table a had been formed from hunk h', if only table a had been formed from hunk h'' instead of from hunk h. Take care here. Since a could not have been formed from h', it is arguable that any proposition, and hence any materially complete proposition, would have been true if a had been formed from h', or alternatively, that no proposition, and hence no materially complete proposition, would have been true if a had been formed from h'. Though it is not *in fact* possible for a to have been formed from h', it *might have been* possible, and indeed it would have been possible if only a had been formed from h''. Proposition p' is a (the) materially complete proposition such that: if a had been formed from h'', then it would have been the case that if a had been formed from h', then p' would have been true. Then we have:

---

world identity and distinctness—are metaphysically "grounded in," and "consist in," nonidentity facts about x and y, so that such identity facts do not obtain independently and solely by their own hook but only *in virtue of* nonidentity facts.

Of course, this formulation does not make clear the exact import of the intended principle. Forbes's intent can be gleaned to a certain extent by noting what he takes the principle to entail. An argument purporting to disprove the principle (whatever its precise import) is given in Section 46.

(C_4) $\quad\Diamond[p' \,\&\, {\sim}M(a, h')] \,\&\, \Diamond\Diamond[p' \,\&\, M(a, h')] \,\&\, {\sim}\Diamond[p' \,\&\,$ $M(a, h')]$.

More intuitively there is a world w accessible to the actual world in which a table distinct from a is formed from hunk h'. There is also a world w' accessible to some world accessible to the actual world (through none accessible to the actual world itself) that is exactly like w in every detail concerning the very matter it contains, with its exact configuration and causal interconnections throughout time, atom for atom, quark for quark, but in which a is the table formed from hunk h'. Worlds w and w' are materially, and hence also purely qualitatively, indistinguishable. Exactly the same material facts obtain in both. Though they are materially indiscernible, they differ in their accessibility relations. World w is accessible to the actual world, whereas world w' is not. Hence, by the Indiscernibility of Identicals, the two worlds are distinct.

An unbridled principle of the identity of materially indiscernible worlds is refuted by the example of the worlds w and w'. Though materially indiscernible, the worlds w and w' are indeed discernible, and not merely by their accessibility relations to the actual world. They also differ as regards which facts obtain in them. World w' includes the fact that a is the table formed from hunk h', whereas world w excludes this. In w, some table distinct from a is the table formed from hunk h'. It follows again by the Indiscernibility of Identicals that the worlds w and w' are distinct.

The temptation to identify the worlds w and w' may stem, in part, from misconceiving possible worlds as material objects, or as entities made solely of matter.[26] Possible worlds are abstract entities

---

[26]This is not the only likely source of the temptation. Another possible source stems from the natural and plausible reductionist principle that a table is "nothing over and above" its matter, in the sense that a complete accounting of all of the matter in a genuinely possible world, with its exact configuration throughout time, must determine all of the remaining facts about the material objects, like tables, and everything else, present in the world. This immediately yields the supervenience thesis mentioned at the end of Section

whose structure comes from the facts that obtain in them. We saw in Section 41 that worlds may be conceived as maximal (in the weak sense) consistent sets of propositions, or total (in the weak sense) histories or states of the cosmos, or maximal states of affairs, or total scenarios, and so on. Consider the first conception: worlds as maximal consistent sets of propositions. Then w and w' are maximal consistent sets that both include the materially complete proposition p' as an element. The set w' includes the further proposition that the table formed from hunk h' is a, whereas the set w includes the further proposition that the table formed from hunk h' is some table distinct from a. Both sets are maximal consistent. Thus both are equally legitimate as worlds *per se*. Though they are not disjoint, they are unquestionably distinct sets.

Similar remarks may be made with respect to any of the alternative conceptions of the worlds w and w'. In fact, these various conceptions of worlds strongly suggest an alternative to simple material indiscernibility as a criterion for identity between worlds. They suggest a principle of the identity of *factually* indiscernible worlds, worlds in which the very same facts obtain. One might also endorse an independent principle of the identity of mutually accessible materially indiscernible worlds (a version of the supervenience thesis mentioned at the end of Section 36 above) or a principle of the identity of materially indiscernible worlds accessible to the actual world. On any of the conceptions of worlds mentioned here, an unrestricted principle of the identity of simply materially indiscernible worlds is straightforwardly false.

---

36 in connection with the fallaciously obtained false conclusion ($C2$). Within an $S4$ framework, the reductionist principle renders the worlds w and w' exactly alike in all of the facts that obtain in each and in their accessibility relations. Cf. this volume, pp. 237–238.

It should be noted that the conclusion (C_4), which I claim to be true, is inconsistent in $S4$ modal logic.

45. Necessity and Apriority

Forbes has raised a second sort of objection to the intransitive-accessibility solution to the modal paradoxes. He argues that if we consider essentialist principles like (III), "we see that there is a conceptual character to such claims," and that metaphysical necessity is "fundamentally an *a priori* matter, to do with the content of our concepts [for example, our concepts of a table and of original matter], even though with the addition of *a posteriori* information, necessary *a posteriori* truths can be inferred."[27] Furthermore, "any *a posteriori* truth *p* necessary at the actual world is so by being true at the actual world and by some conceptual [*a priori*] truth's entailing that *p*'s truth makes it necessary."[28] Since metaphysical necessity is thus the product of conceptual apriority, Forbes argues, every instance of the *S4* axiom schema is indeed true. For if it is conceptually *a priori*, and consequently necessary, that *p*, then it is also conceptually *a priori* that it is conceptually *a priori* that *p*. And if it is necessary but *a posteriori* that *p*, then it is nevertheless conceptually *a priori*, and consequently necessary, that if *p* then it is necessary that *p*. From this it follows (in even the weak system *T* of modal propositional logic) that if it is necessary but *a posteriori* that *p*, then it is still necessary that it is necessary that *p*.

It may be true that conceptual apriority entails metaphysical necessity, in the sense that (with somewhat rare, and for present purposes irrelevant, exceptions) anything that is conceptually *a priori* is generally *ipso facto* metaphysically necessary. Probably something like this accounts for the fact that (II) is not only necessarily true, but it is also necessary that it is necessarily true, and it is necessary that it is necessary that it is necessarily true, and so on. As Forbes acknowledges in presenting his argument, there are examples—coming primarily from the work of Kripke—of propositions that are

[27]Forbes, 1984, p. 185.
[28]Forbes, 1985, p. 237 n. 26. I have replaced Forbes's letter 'T' by 'p' for perspicuity.

metaphysically necessary yet conceptually *a posteriori*. With respect to these examples, the argument that *a priori* necessity iterates—the argument that if it is necessary, because *a priori*, that p, then it is also necessary that it is necessary that p, and so on—is inapplicable. The argument is inapplicable precisely because the examples in question, though necessary, are not *a priori* and hence not necessary-by-virtue-of-being-*a-priori*.

The propositions that the intransitive-accessibility account holds to be necessary but not doubly necessary (for example, the proposition that table a is not originally formed from hunk h_m) are precisely certain *a posteriori* propositions whose necessity is derived by means of *a priori* modal principles like (III) taken together with certain further information, at least some of which is not *a priori*. That is, the propositions that the intransitive-accessibility account holds to be necessary but not doubly necessary are propositions of precisely the sort that Kripke cites as necessary yet a posteriori. The *a priori* principle (III) might be used to establish the necessity of table a's not originating from hunk h_m, but the fact that a does not thus originate is itself unquestionably empirical and not *a priori*.

In fact, not even the conditional 'If table a is not originally formed from hunk h_m, then it is necessary that a is not originally formed from h_m' is *a priori*. For all that is known *a priori*, table a may have originated from hunk h_1, in which case a would still not have originated from hunk h_m, although it would then be possible for a to have thus originated.[29] The necessary *a posteriori* truth that table a is not formed from hunk h_m is thus a counterexample to Forbes's claim concerning the source of necessary *a posteriori* truths. Since the conditional proposition that if a is not formed from h_m then a is necessarily not thus formed is not *a priori*, it cannot be entailed by any conceptual *a priori* truth.

That a's not originating from h_m is in fact necessary yields no

---

[29]The epistemological status of such propositions as these is discussed further in Section 46.

reason to suppose that it must also be doubly necessary, triply necessary, and so on. Indeed, the fact a does not originate from h_m might not have been necessary at all.

The accessibility account rejects the *S4* axiom in its unrestricted form, but the account allows that there may be interesting special cases of necessity iteration that are logically valid. For example, it may be that, as Forbes's argument suggests, necessity iteration is legitimate whenever the proposition in question is necessary by virtue of being *a priori*. Certainly necessity iteration is legitimate with respect to purely mathematical propositions and (classical) logical truths. Maybe here is a legitimate restricted version of the *S4* axiom schema, or the rule of necessity iteration. Are there others? Necessity iteration is fallacious with respect to certain *a posteriori* propositions, but are there any necessary *a posteriori* propositions with respect to which necessity iteration is a legitimate logical inference? For example, Kripke and Putnam have argued that it is necessary, even though *a posteriori*, that cats are animals, or at least that cats are not robots. Presumably, they would argue that it is even necessary that it is necessary that cats are not robots. Does the latter modal fact follow logically from the former, taken together with certain information concerning the nature of the proposition that cats are not robots? If so, what is it about the proposition that cats are not robots that allows for necessity iteration as a logical inference, whereas necessity iteration with respect to other *a posteriori* propositions is fallacious?

The questions raised here seem to be worthy of further research. These and other challenging philosophical questions arise directly from the modal paradoxes. This alone makes the paradoxes deserving of our attention.

46. The Determinacy of Identity

The proof that identity is nonvague and either determinately true or determinately false for any pair of objects of any kind whatsoever proceeds from the observation that if there is a pair of objects, x and y, of which the 'is' of identity is neither determinately true nor determinately false (i.e., there is no objective, determinate fact of the matter whether x and y are numerically identical), then since the 'is' of identity is absolutely determinately true of the pair $<x, x>$, the two pairs must be different pairs of objects. It follows that the objects x and y are themselves distinct. In that case, the 'is' of identity is, contrary to the hypothesis, defined as determinately false for the pair $<x, y>$. Therefore, there is no pair of objects of any kind for which the question of their identity is metaphysically indeterminate.[30]

Although this proof is very convincing—in fact, to my mind, conclusive—I have found that (like most arguments against firmly entrenched philosophical views) it does not always convince. By far the most common objections I have encountered are based on the contention that the proof relies on principles of classical reasoning, whereas the view it purports to refute demands some special nonclassical logic of vagueness. Hence, it is worth emphasizing that the proof does not illegitimately assume or presuppose classical two-valued logic. To assume that every identity proposition is either true or false would certainly be eristically illegitimate, since the argument is advanced against a view that requires a nonclassical, nonbivalent logic. The critical move in the proof is a simple Leibniz's Law inference from an assumption of the form ⌜α has a property F that β does not have⌝ to its trivial consequence ⌜$\alpha \neq \beta$⌝. Even on the view

---

[30]The proof is elaborated and defended on pp. 243–245 of this volume. The general form of the argument was first given by Gareth Evans (1978), p. 208, although the argument had occurred to me independently. For further discussion, see Broome, 1984; Noonan, 1982, pp. 3–6, and 1984; Over, 1984, p. 6; Thomason, 1982; and Appendix III of this volume, p. 269.

being disputed, any inference from something assumed to be true is legitimate if the inference pattern is such as to preserve truth (or such as to preserve "determinate truth," or "complete truth," or "truth to the maximum degree," and so on). Analogously, the term 'bald' (in the sense of 'nearly absolutely bald') is unquestionably vague, in that there are (or at least there could be) individuals who have very little hair on their heads but just enough so that it is neither true nor false (vague, indeterminate, there is no objective fact of the matter) that they are bald. It would be illegitimate to assume that every proposition concerning whether someone is bald is either true or false. A nonclassical, nonbivalent logic is needed in order to reason properly with respect to such propositions. Despite this feature of the term 'bald', if one assumes for the sake of argument (e.g., for a *reductio* argument or for a conditional proof) that Harry has a full head of hair, it is perfectly legitimate to infer that Harry is not bald, for we have assumed as determinately true information that is such that if it is determinately true, then so is the proposition that Harry is not bald. There can be no question but that the Leibniz's Law inference invoked in the proof of the determinacy of identity is likewise such as to preserve determinate truth and is therefore likewise legitimate. Whatever x and y may be, they are not one thing if they differ in any way, since any one thing has every property it has. Nothing could be more trivial.[31]

The critical premise involved in the proof is the assumption that the 'is' of identity is determinately true of any object and itself, and determinately false of any pair of determinately distinct objects. Lest anyone wish to challenge this assumption, it is important to recall Kripke's powerful '*schmidentity*' method of philosophical argument (which ironically applies virtually unchanged to the present case):[32] We may invent a new sense of 'is'—the 'is' of schmidentity—such that our assumption is true solely by stipulation for this new sense.

[31]Cf. this volume, p. 244n.
[32]Kripke, 1972, p. 108.

Then we may prove, by the now familiar argument, that schmidentity is fully defined and determinate for any pair of objects. Yet this allegedly new sense of 'is' is precisely one that gives rise to the very sorts of problems for which the theory of indeterminate identity was introduced in the first place. Who cares about any other alleged sense of 'is' when our concern is with a question of identity? What's so important about x and y being neither determinately "identical" nor determinately "distinct" in some other sense if they are determinately not one and the very same but two, determinately not schmidentical? Nothing. Where one's concern is with a question of numerical identity, *almost* doesn't count. In fact, it doesn't even make sense.[33]

There is an alternative way of constructing the proof, one which applies the Leibniz's Law inference directly to the objects x and y: Suppose again that it is indeterminate whether x and y are identical. Then x and y differ in that x is determinately identical with x, whereas y is not. That is, y does not have x's property of being such that the 'is' of identity is determinately true of the ordered pair of x together with it. Hence, contrary to the hypothesis, x and y are determinately distinct.

This alternative construction reveals that the general form of the argument is essentially that used in proving the necessity of identity as a theorem of quantified modal logic: For every x and every y, if $x = y$, then it is necessary that $x = y$.[34] More analogously, the argument parallels the proof of the contrapositive of the necessity of identity, a theorem that Alonzo Church has called 'Murphy's Law of Modality': For every x and every y, if it is possible that $x \neq y$, then (since y does not have x's property of being necessarily identical with x) $x \neq y$.

Church has recently used this general form of argument to argue that if quantification into propositional attitude contexts is accepted as meaningful together with the usual laws of classical logic, then it

---

[33]Cf. this volume, pp. 244–245.

[34]See Barcan, 1947.

is very likely that for every *x* and every *y*, if someone believes that *x* ≠ *y*, then *x* ≠ *y*.[35] The general argument can also be used to establish—or at least to argue compellingly for—a number of other philosophically interesting and highly controversial (in some cases, nearly universally denied) theses, such as the following:

T1: For every *x* and every *y*, if *x* = *y*, then whenever *x* exists, *x* = *y*;

T2: For every *x* and every *y*, if *x* = *y*, then if one believes anything at all involving *x*, one knows that *x* = *y*;

T3: For every *x* and every *y*, the question of whether *x* = *y* is not a matter of decision, convention, or convenience, nor of elegance, simplicity, or uniformity of theory;

T4: For every *x* and every *y*, if *x* = *y*, then *x* is the only possible individual that could possibly have any metaphysically relevant "claim" or "title" to be *y*;

T5: For every *x* and every *y*, the question of whether *x* = *y* does not turn on any fact concerning anything other than *x* and *y*;

T6: For every *x* and every *y*, if *x* = *y*, then the fact that *x* = *y* does not require any "criteria of identity" for things of *x*'s sort or kind;

T7: For every *x* and every *y*, if *x* = *y*, then the fact that *x* = *y* is not grounded in, or reducible to, qualitative nonidentity facts about *x* and *y* other than *x*'s existence, such as facts concerning material origins, bodily continuity, or memory;

[35]Church concludes from this argument and from the provability of the necessity of identity in quantified modal logic that there are compelling reasons to reject the meaningfulness of quantification into either modal or propositional attitude contexts. See Church, 1988; for a response, see Salmon, 1986b.

T8: For every x and every y, if $x = y$, then the fact that $x = y$ obtains by virtue of x's existence, and not at all by virtue of any other qualitative nonidentity facts about x, and y, such as facts concerning material origins, bodily continuity, or memory;

T9: For every x and every y, if one knows that $x = y$, then one knows this (primarily) solely by logic and by one's acquaintance with x, and not by knowing qualitative nonidentity facts about x and y, such as facts concerning continuity, location, or qualitative persistence or similarity.[36]

Each of these theses is diametrically opposed to the views, theories, or presuppositions of some major segment of the contemporary analytic philosophical community. Much of the literature on cross-time identity (and especially on personal identity), for example, presupposes the opposite of one or more of theses T6, T7, and T8. Many of the most widely held theories in this literature involve denying several (and in some cases all) of the remaining theses. Nearly the same is true of much of the literature on cross-world identity. In particular, that cross-world and cross-time identity facts are grounded in nonidentity facts is a recurrent theme in Forbes's work.[37] Although it is evidently not widely recognized, each of the theses mentioned is in fact, despite its unpopularity, a virtual consequence of Leibniz's Law together with some trivial feature of the reflexive law of identity. The trick (if there is. any) is to extract the right property of x from the relevant trivial feature of the law (or proposition or fact) that $x = x$.

Consider thesis T7: Whatever x may be, the trivial fact that $x = x$ is not at all grounded in, or reducible to, any facts about x like those concerning x's material origins, x's bodily continuity through time, or x's memory of past experiences. If the fact that $x = x$ is grounded in any other fact about x, it is only grounded in the mere fact that x

---

[36]For a general defense of thesis T9, see Salmon, 1986a.

[37]See, for example, Forbes 1980, 1983, and 1985, pp. 126–131 and passim.

exists. Thus x has the complex property of being such that the fact that x is identical with it is not grounded in any qualitative nonidentity facts about x other than x's existence. Hence, by Leibniz's Law, for every y, if x and y are one and the very same, then y also has this complex property. Thus, if $x = y$ then the fact that $x = y$ is not grounded in any qualitative nonidentity facts about x (which are also facts about y) other than x's existence. Indeed, since x and y are one and the very same, the fact that $x = y$ is just the fact that $x = x$. Consequently, the fact that $x = y$ must have the property of the fact that $x = x$ that it is not grounded in any qualitative nonidentity facts about x (which are also facts about y) other than x's existence—*QED*. What a trivial and yet wonderful thing is Leibniz's Law!

The original proof that identity is determinate and fully defined for every pair of objects incidentally yields a persuasive reason for believing that the threshold for the amount of different original matter possible in the construction of an artifact consists in a sharp cutoff point rather than in a range of indeterminacy. A simple thought experiment shows that the threshold for the amount of different matter possible in the *reconstruction* of an artifact at some time after its disassembly does indeed consist in a sharp cutoff point. Recall that the number of molecules in the original matter of table a is n. Suppose that at time t_1, n distinct tables, a_1, a_2, ..., a_n, each qualitatively identical to a, are constructed from exactly n molecules apiece. At a later time t_2, each table is completely dismantled. At a still later time t_3, n tables, a_1', a_2' ..., a_n', are constructed according to the same plan in the following way: a_1' is formed from all of the original molecules of a_1 except for the replacement of one molecule by a qualitative duplicate; a_2' is formed from all of the original molecules of a_2 except for the replacement of two molecules by qualitative duplicates of each; and so on, up to a_n', which is formed from entirely new matter. Clearly $a_1 = a_1'$ whereas $a \neq a_n'$. In fact, the construction of the second sequence of tables may be such that, for any i, if $a = a_i'$, then $a_{i-1} = a_{i-1}'$, and if $a_i \neq a_i'$, then $a_{i+1} \neq a_{i+1}'$. By the proof of the

343

determinacy of identity, for any i, it is either determinately true that $a_i = a_i'$, or else it is determinately true that $a_i \neq a_i'$. Therefore, there must be some precise amount of different matter that first passes the threshold for the amount of different matter possible in the reconstruction of table a, i.e., there must be some m such that $a_m \neq a_m'$ but $a_{m-1} = a_{m-1}'$. On certain natural assumptions, this yields an excellent reason for supposing that in the sequence of hunks of matter h_1, h_2, ..., h_n, there is a hunk h_m that is the first to pass the threshold for the amount of different original matter possible in the construction of table a, i.e., that this threshold also consists in a precise cutoff point.

It was noted in Section 45 above that it is *a posteriori*, even though it is necessary, that table a is not originally formed from hunk of matter h_m. Hence, it cannot be *a priori* that a is *necessarily* not originally formed from h_m. Is it then *a posteriori*? It is difficult to imagine establishing, by philosophical argument or otherwise, exactly what number m is, i.e., precisely how many molecules of difference from the actual original matter of table a would first result in a new and different table. It seems likely that it is *unknowable* that table a is necessarily not originally formed from hunk of matter h_m. That is, although it is knowable *a posteriori* that a is not in fact originally formed from h_m, and it is knowable (perhaps even knowable *a priori*) that there is some number m such that a difference of original matter of fewer than m molecules would still result in the same table though a difference of m or greater would result in a different table, it seems unlikely that one could know (*a priori* or *a posteriori*) of the relevant number m, whatever it is, that it is the threshold number of molecules of difference for the potential construction of table a. Whatever number m is, the fact that a is necessarily not originally formed from precisely that many different molecules from a's actual original matter would appear to be a fact that is neither knowable *a priori* nor knowable *a posteriori*, since it appears not to be knowable at all.[38]

---

[38]The present essay has benefited from the helpful comments and suggestions of Pascal Engel, Graeme Forbes, David Lewis, and John Pollock.

Philosophical Review

THE LOGIC OF WHAT MIGHT HAVE BEEN*
Nathan Salmon

In earlier work I argued (following Hugh Chandler) that the conventionally accepted system *S5* of (first-order) modal propositional logic, and even the weaker system *S4*, embody an invalid pattern of modal reasoning; they are fallacious systems for reasoning about what might have been.[1] I argued, in fact, that the characteristic *S4* axiom schema, "□φ⊃□□φ"—or equivalently, the principle that for any necessarily true proposition *p*, the proposition that *p* is necessarily true is itself necessarily true—is not only not logically true, some instances are in fact untrue. I argued, that is, that for some necessary truths *p*—for example, that a certain table does not originate from a certain hunk of wood—the fact that *p* is necessary cannot itself be correctly deemed necessary. Instead, although any such proposition *p* is necessary, the claim that *p* is necessarily necessary is untrue, and indeed some claim of the form "□□. . .□*p*" is altogether false.

While some of my audience have found these arguments against *S4* modal logic persuasive, many have found them unconvincing. I have repeatedly encountered two particular objections, which are probably best regarded as two parts of a single objection. This objection, however, betrays a serious misunderstanding of my position, or a failure to appreciate the full force of my (Chandleresque) arguments, or both, and is based on a confusion among concepts central to the foundations of contemporary semantics for

---

*This paper was presented to an international conference on Meaning and Natural Kinds at the Inter-University Centre of Postgraduate Studies in Dubrovnik, Yugoslavia, September 1986. It has benefitted from a discussion with Timothy Williamson, and from comments by Hugh Chandler, Graeme Forbes, and the anonymous referees.

[1] Hugh Chandler, "Plantinga and the Contingently Possible," *Analysis* 36 (1976), pp. 106–109. For my renderings of Chandleresque arguments, see *Reference and Essence* (Princeton, N.J.: Princeton University Press and Oxford, England: Basil Blackwell, 1981), séction 28, pp. 229–252; "Impossible Worlds," *Analysis* 44 (1984), pp. 114–117; and "Modal Paradox: Parts and Counterparts, Points and Counterpoints," in French, Uehling, and Wettstein, eds., *Midwest Studies in Philosophy XI: Studies in Essentialism* (Minneapolis, Minn.: University of Minnesota Press, 1986), pp. 75–120. The last includes further bibliographical references.

modal logic. In this paper I shall present the objection(s) and my response. I shall also argue for the further claim (which I have not made elsewhere) that even the conventionally accepted system B, which is weaker than $S5$ and independent of $S4$, has not been adequately justified as a fallacy-free system of reasoning about what might have been. The axioms characteristic of B are sentences of the form "$\phi \supset \Box \Diamond \phi$." That is, B is characterized by the principle that for any true proposition p, the proposition that p is possibly true is itself necessarily true. Here, however, I shall not argue for the strong claim (analogous to my claim in connection with $S4$) that some true proposition p is such that the proposition that p is necessarily possible is untrue. (I believe that the characteristic B principle may well have no such counterexamples.) I contend only that, even if the B axioms are in fact true, and even if they are necessarily true, it seems to be logically possible for some proposition p to be true while the proposition that p is necessarily possible is at the same time false. Thus, even if the B principle is necessarily true, its alleged status as a logical (or analytic) truth remains in need of justification. Similar arguments may be made against other proposed extensions of the weak modal system T. If I am correct, insofar as modal logic is concerned exclusively with the logic of metaphysical modality, and not also with other, nonlogical features of metaphysical modality, T may well be the one and only (strongest) correct system of (first-order) propositional modal logic.[2]

I.

The case against $S4$ modal logic stems from the intuition (which

---

[2]Metaphysical modal logic concerns metaphysical (or alethic) necessity and metaphysical (alethic) possibility, or necessity and possibility *tout court* —as opposed to such other types of modality as physical necessity, epistemic necessity, etc. The (strongest) correct system of logic for some other modality need not coincide with that for metaphysical modality. (The characteristic principle of T that any proposition that must be true is true must already fail in deontic modal logic, the logic of what is morally required to be the case and what is morally permitted to be the case.) Throughout this paper I am concerned primarily with metaphysical modality. Where I speak simply of "modal logic," the reader is to understand that only metaphysical modal logic is under discussion. My use of such modal locutions as "necessary," "might have," etc. is to be similarly construed throughout, unless otherwise indicated.

4

many of my opponents share) that a particular material artifact—
say, a particular wooden table which we may call "Woody"—could
have originated from matter slightly different from its actual origi-
nal matter m^* (while retaining its numerical identity, or its *haecceity*)
but not from entirely different matter. Wherever one may choose
to draw the line between what matter Woody might have origi-
nated from and what matter Woody could not have originated
from, it would seem that, by stretching things to the limit, we may
select some (presumably scattered) matter m such that, although
Woody could not have originated from m, m is close enough to
being a possibility for Woody that if Woody *had* originated from
certain matter m' that *is in fact* possible for Woody—matter dif-
fering in as many molecules from the actual original matter m^* as
possible, and sharing as many molecules with m as possible, while
remaining a possibility for Woody—then it *would have been* possible
for Woody to have originated from m, even though it is not *actually*
possible. Even if one denies that there is a sharp line to be drawn
between what matter is and what matter is not possible for the
origin of Woody, by stretching things to whatever sort of limit re-
mains (such as an interval of vagueness and indeterminacy in lieu
of a dividing line between what is and is not possible), there will
still be some matter m such that Woody (just barely) determinately
could not have originated from m, yet the claim that this is itself
necessary is untrue (or not "true to the maximum degree," or
whatever), and in addition, unfalse. Either way, the conditional
claim (which is an axiom of *S4*) that if Woody necessarily does not
originate from m, then it is necessary that Woody necessarily does
not thus originate fails. (It suffers the same truth-value status as its
consequent.) Also failing is the inference from the antecedent of
this conditional to the consequent, since the premise of the infer-
ence is altogether true and the conclusion is not. *S4* modal logic is
fallacious.

I supplemented my argument against *S4* with a particular con-
ception of what possible worlds are—in conjunction with the stan-
dard identification of necessity with truth in every possible world
and possibility with truth in at least one possible world. As with
many contemporary philosophers of modality, I conceive of pos-
sible worlds as certain sorts of (in some sense) maximal abstract
entities *according to which* certain things (facts, states of affairs) ob-
tain and certain other such things do not obtain. Possible worlds

5

are total *ways things might have been* (David Lewis). A possible world is something like a total history that might have obtained concerning everything in the cosmos (Saul Kripke), or a maximal property or state that the cosmos might have had or been in (Kripke, Robert Stalnaker), or a maximal state of affairs (Alvin Plantinga) or maximal scenario (myself) that might have obtained. For most purposes, one may conceive of a possible world as an infinitely long, complex, and detailed set of states of affairs or (potential) facts or statements (that is, an infinite set of structured propositions, more or less as Russell conceived propositions[3]), one that does not leave any question of fact undecided (Robert Adams). Since the actual world is itself a possible world, it too is conceived of as a maximal scenario or history, and may be conceived of as a maximally comprehensive set of statements, in this case the set of all statements that are in fact true.

More accurately, a possible world may be conceived of as a set of (potential) facts or statements that does not leave any of a very comprehensive range of questions of fact undecided. Some of the facts that are decided may in some cases determine that certain other statements are neither true nor false, owing to false presuppositions, category mistakes ("sortal incorrectness"), vagueness, or something else. If Frege was right, for example, the fact that there is no present King of France determines that the statement that the present King of France is bald is neither true nor false, so that neither this statement nor its negation is included in the set of statements corresponding to the actual world. More importantly, certain meta-facts (or facts about possible worlds and sets of facts) cannot be included in such a set for familiar reasons concerning cardinality problems, since there are at least as many such meta-facts as there are subsets of any given infinite set of facts, and these subsets outnumber the facts in the given set.[4] A possible world,

---

[3]I mean to exclude here the modal logician's conception of a proposition as a set of possible worlds (or equivalently, as a characteristic function from possible worlds to truth values). It is not a good idea to think of possible worlds as sets of propositions, and at the same time to think of propositions as sets of possible worlds. For more on my favored Russellian conception of propositions, see my *Frege's Puzzle* (Cambridge, Mass.: The MIT Press/Bradford Books, 1986).

[4]See Selmer Bringsjord, "Are There Set Theoretic Possible Worlds?" *Analysis* 45 (1985), p. 64; Christopher Menzel, "On Set Theoretic Possible Worlds," *Analysis* 46 (1986), pp. 68–72; and Patrick Grim, "On Sets and Worlds: A Reply to Menzel," *Analysis* 46 (1986), pp. 186–191.

then, may be thought of as a set of statements of a certain restricted but still very comprehensive sort.

Recall that it is (just barely) impossible for Woody the table to have originated from certain matter m. Woody cannot be in the state of originating from m. That is, originating from m is a state metaphysically unavailable to Woody; it is a way that Woody cannot be. But it is still a way for an individual to be. Likewise, there is a total way for all things in general to be—a "maximal" set of (potential) facts, if you will—according to which Woody originates from m. Let us call this maximal way for things to be "W." Since Woody originates from m according to W, and Woody metaphysically cannot do so, W is a total way things cannot be. A total way things cannot be is a total way for things to be such that things cannot be that way, a state or history for everything in the universe such that everything in the universe cannot be in that state or have that history, a maximal state of affairs or scenario that cannot obtain. Total ways things cannot be are thus also "worlds," or maximal ways for things to be. They are impossible worlds. In fact, although W is an impossible world, there is a possible world W' (assuming m was chosen carefully enough, and ignoring for the moment the prospect of vagueness and regions of indeterminacy) according to which Woody originates from the matter m' instead of its actual original matter m^*, and if W' had obtained (as indeed it might have), W would have been a way things might have been rather than a way things cannot be; W would have been possible instead of impossible. Although W is impossible relative to the actual world, it is possible relative to W', which is itself possible relative to the actual world. Thus W is a possibly possible world. Other impossible worlds may be not even possibly possible, but only possibly possibly possible, and so on. The binary relation between (possible or impossible) worlds of relative possibility—the modal relation of *accessibility*— is not transitive.

What are the limits on the admissibility of possible and impossible worlds? None to speak of. Any degree of variation and recombination qualifies. Some ways for things to be are not even possibly possibly . . . possible, for any degree of nesting. A world according to which Nathan Salmon is Henry Kissinger is such a world, for example, as is a world according to which Nathan Salmon is a Visa credit card account with the Bank of America. Since they are ways-for-things-to-be of a certain sort (viz., such that

7

things necessarily cannot be that way, and necessarily necessarily cannot be that way, and so on), these too are "worlds." As far as I can tell, worlds need not even be logically consistent. A world according to which there is both life on Mars and no life on Mars is a way things cannot be on logical grounds alone. Hence this too is a "world," a way for things to be. The only restriction on *worlds*, as opposed to lesser ways for things to be, is that they must be (in some sense) *maximal* (total, comprehensive) ways for things to be; for every statement of fact, either it or its denial must obtain according to a world—*modulo* cases of nonbivalence arising from presupposition failure, vagueness, etc., and subject to cardinality constraints if the totality of facts comprising a world are to form a set.

II.

The first part of the standard objection to this account is summed up by David Lewis as follows:

> Say I: This is no defence [of the essentialist doctrine that a table could not have originated from entirely different matter], this is capitulation [to radical anti-essentialism]. In these questions of haecceitism and essence, by what right do we ignore worlds that are deemed inaccessible? Accessible or not, they're still worlds. We still believe in them. Why don't they count? (*On the Plurality of Worlds*, Oxford: Basil Blackwell, 1986, p. 246).

This part of the objection may be spelled out further: Intransitive accessibility relations are introduced into modal semantics for the purpose of interpreting various "real" or restricted types of modalities, such as nomological necessity. A proposition is *nomologically necessary* in an arbitrary possible world w if and only if it is true in every possible world in which all of the laws of nature in w are true. For convenience, we may say that a world w' is *accessible to*, or *nomologically possible relative to*, a world w if every natural law of w is true in w'. Then we may say more succinctly that a proposition is nomologically necessary with respect to a possible world w if and only if it is true in every possible world accessible to w. More restrictedly, perhaps, a proposition is *physically necessary* with respect to an arbitrary possible world w if and only if it is true in every possible world in which all of the laws of physics in w are

8

true. Other restricted modalities require alternative accessibility relations: a proposition is said to be *necessary*, in the restricted sense in question, with respect to an arbitrary possible world w if and only if it is true in every possible world of such-and-such a restricted sort—the restriction in question depending on some appropriate relation to w. Such restrictions yield failures of the characteristic *S4* principle that any "necessary" truth is necessarily necessary, and even of the characteristic *B* principle that any truth is necessarily possible. Suppose, for example, that w and w' are worlds so different in their natural constitution that although every natural law of w is true in w' (so that w' is nomologically possible relative to w), some of these natural laws of w are not natural laws in w' but merely accidental generalizations, while certain other generalizations not even true in w are additional natural laws in w'. Then a natural law of w (which is automatically nomologically necessary in w) that is not also a natural law of w' will not be true in every world nomologically possible relative to w', and hence will not be nomologically necessarily nomologically necessary in w. Similarly, a proposition that is true in w but violates one of the additional natural laws of w' will not be nomologically necessarily nomologically possible in w. In this restricted scheme, accessibility between worlds is neither transitive nor symmetric. It remains reflexive, of course—as long as the natural laws of a given world are true in that world. The fundamental characteristic T principle that any "necessary" truth is true is thereby preserved.

By contrast, the objection goes, the hallmark of metaphysical (alethic) necessity or necessity *tout court*—its distinguishing characteristic—is that it is completely unrestricted. Metaphysical necessity and possibility is the limiting case of restricted necessity and possibility, the case with no restrictions whatsoever. A proposition is necessary in this unrestricted sense with respect to a possible world w if and only if it is true in absolutely every possible world whatsoever, no restrictions. By contrast with the case of restricted modalities, the objection continues, my conception of a *metaphysically impossible world* is incoherent. Any possible world is *possible* in the unrestricted, metaphysical sense. Since my account admits the existence of a world W in which Woody originates from m, even though I deem this world "inaccessible" to the actual world, I implicitly acknowledge (contrary to my explicit pronouncements) that it is *not* necessary in the relevant, metaphysical sense of "neces-

sary" that Woody does not originate from m. Indeed, by admitting possible worlds of unlimited variation and recombination, I simply abandon true metaphysical essentialism. By my lights, any property is attached to anything in some possible world or other. I am a closet radical anti-essentialist.

This part of the objection brings with it an oft-used defense of $S5$ modal logic. In the metaphysical, unrestricted senses of "necessary" and "possible," the characteristic $S5$ principle that any possible truth is necessarily possible may be easily proved. Suppose p is a possible truth, that is, a proposition true in at least one possible world w. Then relative to any possible world w', without exception, there is at least one possible world in which p is true—namely, w. It follows (given our assumption that p is possible) that it is necessary that p is possible. For in the unrestricted sense of "possible," one possible world in which p is true is all that is required for p to be "possible" relative to any given world w', with no further restriction as to what sort of world p is true in or how that world is related to w'. There are similar direct proofs of the characteristic B and $S4$ principles.

There remains my claim that such a world as W, in which Woody originates from m, is inaccessible to the actual world. The first part of the objection more or less ignores this claim as irrelevant, a red herring. The second part of the objection focuses on this claim. When such restricted modalities as nomological necessity or physical necessity are under discussion, the phrase "possible relative to" has a tolerably clear sense (given that we have a prior understanding of such notions as *law of nature* and *law of physics*). Such notions of accessibility are more or less sharply defined. My notion of *necessity* is also some restricted notion, since I deem some worlds inaccessible to others. Yet, the objection goes, I have not defined the restriction; I leave my use of the phrase "possible relative to" with no tolerably clear sense. It does not seem to mean much of anything; it is simply an *ad hoc* device for sweeping a serious difficulty under the rug. To quote Lewis again:

> [W]e look in vain, in . . . many . . . places, for an account of what it means to deny that some world is 'relatively possible'. I think it is like saying: there are things such that, ignoring them, there are no such things. Ignoring all the worlds where such-and-such obnoxious things happen, it is impossible that such things happen. Yes. Small comfort (ibid., p. 248).

10

III.

The objection presented in the preceding section confuses or conflates two notions that must be kept sharply distinct: the generic notion of a way for things to be and the peculiarly modal notion of a way things *might have* been. Confusion between these two notions probably stems from an analogous ambiguity in the phrase "possible world." The layman speaks of a "world" almost exclusively as a planet, though sometimes as the whole physical universe of atoms, molecules, planets, stars, galaxies, superclusters, and what-have-you. By contrast, in the metaphysics of modality a *world* is an abstract entity according to which some things obtain and other things do not, such that all (or sort of all) such questions of fact are answered one way or the other. Modal worlds are not physical universes but intensional entities that represent things as being one way or another. Even Lewis, who in his metaphysical constructions idiosyncratically maintains the layman's conception of a *world* as a whole physical universe, combines this conception with the metaphysician's conception of a world as an entity *according to which* some states of affairs obtain (including, for Lewis, states of affairs concerning things not part of that world) and other such states of affairs do not, such that all (or sort of all) such questions of fact are answered by the "world."[5] It is awkward to call these things simply "worlds," since that term is so highly suggestive of the layman's notion. Fortunately (or rather unfortunately!) Leibniz provided a more descriptive term: "possible world."

There are two problems with this bit of Leibnizian terminology. The first problem concerns what the word "possible," as it occurs in the phrase "possible world," does not mean. In metaphysics when we call something a possible such-and-such, we generally mean that it is a such-and-such that might have existed, even if it does not. But whether or not possible worlds actually exist, in calling something a "possible world" most of us do not mean a world (*qua* total way for things to be, or maximal entity according to which some states of affairs obtain and others do not) that might

---

[5] I criticize Lewis's views concerning the nature of possible worlds in my review of his *On the Plurality of Worlds*, in *The Philosophical Review* 97 (1988), pp. 237–244.

have existed, even if it does not. To think that the concept of a possible world is that of a world that might have existed is to misunderstand the function of the word "possible" in the phrase "possible world."[6]

The second problem with the phrase "possible world" concerns what the word "possible" does mean there. For it means *something* there. Strictly speaking, a possible world is not a way for things to be that might have existed; it is a way for things to be such that things might have been that way. Similarly, a possible history or possible state for an individual is not a history or state that might have existed, but a history or state that the individual might have had or might have been in. Thus the word "possible" contributes some special meaning to the phrase, and more meaning than is accommodated by the generic notion of a total way-for-things-to-be-even-if-things-could-not-have-been-that-way. Strictly speaking, a possible world is not any old total way for things to be, but a modally special kind of total way for things to be, namely a total way that things *might have* been. A possible world is a total way for things to be that conforms to metaphysical constraints concerning what might have been. The generic notion of a total way for things to be is a notion without a proper term of its own. Aesthetic considerations aside, rather than let the phrase "possible world" do double duty for this generic notion as well as for the modal notion, we would be better off reserving it exclusively for the modal notion—for which it is certainly more apt—and using my highfalutin hyphenated phrase "total way-for-things-to-be-even-if-things-could-not-have-been-that-way" for the generic notion, or my modally unadorned phrase "total way for things to be," or if worse comes to worst, the simple unadorned word "world." In the best of all possible worlds, total ways for things to be are not called "possible worlds," unless they are total ways things might have been.

---

[6]The objection of the preceding section need not depend in any way on this common misconstrual of the phrase "possible world," although it probably often does. One who misunderstands the phrase "possible world" to mean *world that might have existed* will conclude that "impossible worlds" cannot exist. Possible worlds would emerge as the only worlds there could be, so that a (possible) thing is a world if and only if it is a "possible world." It seems likely that this fallacy lies behind the common confusion of the generic notion of a world and the modal notion of a possible world.

Whatever the source of the confusion between the generic no-
tion of a way for things to be and the modal notion of a way things
might have been, this confusion is very probably the primary
source of the idea that metaphysical modality is the limiting case of
restricted modalities, that metaphysical necessity and possibility is
the unrestricted, and hence the least restricted, type of necessity
and possibility. For metaphysical necessity is indeed truth in all
ways things might have been (modal, not generic), and metaphysi-
cal possibility is indeed truth in at least one way things might have
been (modal, not generic).

Metaphysical modality is definitely *not* an unrestricted limiting
case. There are more modalities in Plato's heaven than are dreamt
of in my critics' philosophy, and some of these are even less restric-
tive than metaphysical modality. One less restrictive type of mo-
dality is provided by *mathematical necessity* and *mathematical possi-
bility*. A proposition is mathematically necessary if its truth is re-
quired by the laws of mathematics alone, and mathematically
possible if its truth is not precluded by the laws of mathematics
alone. Many metaphysical impossibilities are mathematically pos-
sible, for example, Nathan Salmon being a Visa credit card ac-
count with the Bank of America. Another type of modality less
restrictive than metaphysical modality is provided by what is some-
times called "logical necessity" and "logical possibility," to be dis-
tinguished from genuinely metaphysical necessity and possibility,
or necessity and possibility *tout court*. A proposition is logically nec-
essary if its truth is required on logical grounds alone, logically
possible if its truth is not ruled out by logic alone (that is, if its
negation is not logically necessary). Thus whereas it is logically nec-
essary that Nathan Salmon is not somebody other than Nathan
Salmon, and it is also logically necessary that either Nathan Salmon
is a Visa credit card account with the Bank of America or he is not,
it is not logically necessary that Nathan Salmon is not a credit card
account. Although there is a way things logically could be ac-
cording to which I am a credit card account, there is no way things
metaphysically might have been according to which I am a credit
card account. This illustrates the restricted nature of metaphysical
modality. Some logically possible worlds must be "ignored." Meta-
physical necessity is truth in every logically possible world of a cer-
tain restricted sort.

What is the restriction? To worlds that are metaphysically pos-

13

NATHAN SALMON

sible. (What else!) When we identify necessity with truth in every possible world, the word "possible" means *something* there, and what it means there places a restriction on the sort of worlds under consideration. The metaphysical notion of possibility restricts the logical notion of possibility, in a manner exactly analogous to that in which the notion of *natural law* involved in the notion of nomological necessity restricts the metaphysical notion of possibility. Just as nomological possibility is a special kind of metaphysical possibility, so metaphysical possibility is a special kind of logical possibility.[7]

Even logical necessity may be seen as observing some restriction: a proposition is logically necessary (with respect to a world w) if and only if it is true in every *logically consistent* world (according to w), whether metaphysically possible or not—or every world in which the laws and rules of logic (in w) obtain (including the logical prohibition on inconsistency).[8] The logically inconsistent worlds do not count as regards what is logically necessary. Still, logical modality is considerably freer of restriction than metaphysical modality. With its freedom from the additional constraint of metaphysical possibility, logical necessity may be construed as accommodating all of the axioms and rules of *S5*. But if logical modality is unrestrictive enough to accommodate all of the axioms and rules

---

[7]Timothy Williamson has pointed out that this may be strictly false, since (as David Kaplan has shown) there are sentences that are valid in the logic of indexicals and that do not express metaphysically necessary truths, for example "If Saul Kripke is an anthropologist, then Saul Kripke is actually an anthropologist." I believe, however, that insofar as propositions (as opposed to sentences) may be appropriately called "logically valid" or "not logically valid," the propositions expressed by such sentences are not logically valid even though the sentences themselves are. (Conversely, some sentences that are not logically valid express propositions that are, for example, "All bachelors are unmarried men.") Cf. *Frege's Puzzle*, pp. 132–151, and especially p. 177, note 1. The important point here is that some logically possible (that is, consistent) propositions are nevertheless metaphysically impossible.

[8]If w is itself logically consistent, this rules out worlds in which such logical truths as the Law of Noncontradiction do not obtain. What about an inconsistent world according to which there is both life on Mars and no life on Mars and yet (by logic) no proposition and its negation are both true? (I owe this marvelous example to Saul Kripke, who has used it for a different but related purpose.) This had better count somehow as a world in which the Law of Noncontradiction does *not* obtain, in the relevant sense. Otherwise, such contradictions will emerge as logical possibilities.

14

of *S5*, it may not be restrictive enough to zero in on *S5*. Depending on what counts as logically possible, the interpretation of the diamond "◇" as logical possibility instead of metaphysical possibility could turn ⌜◇φ⌝ into a logical truth for every logically consistent formula φ. It would then become a logical truth that Woody "might have" originated from *m*, and that Nathan Salmon "might have" been a credit card account. Even if we essentialists are wrong and metaphysical necessity does not extend beyond logical necessity, the logic of logical necessity can extend far beyond that of metaphysical necessity.[9]

If worlds include ways things metaphysically cannot be in addition to ways things metaphysically might have been, then the idea that metaphysical necessity corresponds to truth in every world whatsoever is flatly mistaken. If worlds include ways things logically cannot be, then no proposition is true according to every world and every proposition is true according to some world. I know of no standard or conventional sense of "possible" on which even the proposition that Nathan Salmon is somebody other than Nathan Salmon is "possible." It is not clear that there would be any interest, other than purely formal interest, in a completely unrestricted notion of modality on which anything is possible and nothing is necessary—and there is not much purely formal in-

---

[9]Thus whereas it is metaphysically impossible on my view for Woody to originate from *m*, it may nevertheless be logically true (and hence logically necessary) that it is logically possible that Woody so originates. Whether the sentence "It is logically possible that Woody originated from *m*" should itself count as a logical truth may depend on whether logical necessity and possibility are treated as attributes of sentences, or rather as attributes of propositional contents. See note 7 above. It is arguable that the logical (as opposed to metaphysical) possibility of truth for the proposition that Woody originated from *m* is itself a truth of pure logic. Alternatively, if logical possibility is an attribute of sentences rather than of their contents, it is arguable that the logic of logical necessity and possibility should take into consideration the logical possibility of the sentence "Woody originates from *m*" being analytically false while retaining its logical form (expressing, for example, the proposition that Venus is distinct from Venus). In that case, it need not be a truth of logic (although it would still be true) that "Woody originated from *m*" is logically possible. Even under this construal, however, *S5* may not be the appropriate (first-order) propositional logic of logical necessity. The rule of necessitation (which licenses the inference of ⌜□φ⌝ from a subsidiary proof of φ) is inapplicable to such logical validities as "If Saul Kripke is an anthropologist, then Saul Kripke is actually an anthropologist." (See note 17 below.)

terest in this unrestricted notion. Such a notion would preserve the characteristic *S4* axiom schema, but perhaps at the cost of turning $\ulcorner \Diamond \phi \urcorner$ into a logical truth for every formula ϕ, and thereby ruling out the inference rule of necessitation (which licenses the inference from a logical theorem ϕ to $\ulcorner \Box \phi \urcorner$) as well as the characteristic axiom schema of *B* and hence also that of *S5*. (The last, in fact, would be replaced by its negation.) Even if there is interest in such a notion, it has nothing to do with metaphysical modality. Surely it is metaphysically impossible that there should be life on Mars and no life on Mars at the same time. The failure of the characteristic *B* axiom schema in the case of the completely unrestricted interpretation of the modal operators demonstrates that there must be some fallacy in the "proof," presented in the preceding section, that unrestricted modality honors *S5*.

Do worlds, *qua* ways for things to be, include ways things cannot be in addition to ways things might have been? I know of no plausible grounds for denying that they do. Indeed, nearly any plausible argument for the existence of ways things might have been (including those arguments offered by my opponents)[10] affords an analogous and parallel argument for ways things cannot be, even ways things cannot be on logical grounds alone. Every argument I am aware of against impossible worlds in favor of only possible worlds confuses ways for things to be with ways things might have been, or worse, confuses ways things cannot be with ways for things to be that cannot exist—or worse yet, commits both errors. The fact that Woody cannot originate from *m* entails that originating from *m* is a way Woody cannot be. It follows from the latter that Woody originating from *m* and Socrates being wise and . . . (where "all" questions of fact are fixed), is a maximal way that things (in general) cannot be. It follows from the fact that Woody cannot originate from *m*, therefore, that there is a maximal way things cannot be. Likewise, it follows from the fact that I cannot be somebody other than myself, that me being somebody other than myself and Socrates being wise, etc., is also a way things cannot be. We should not resist these inferences; we should draw them, and see where they lead. At the very least we should refrain from as-

---

[10]Cf. "Impossible Worlds," pp. 116–117; and Margery Bedford Naylor, "A Note on David Lewis's Realism About Possible Worlds," *Analysis* 46 (1986), pp. 28–29.

16

serting their premises while rejecting their conclusions, since they are valid.

An impossible world like W may be seen as merely a variation of a genuinely possible world. Consider the "maximal" set of statements that would have been true if m had been formed into a table and Woody had never been constructed at all. Let us call the (possible) table that would have been formed from m if m had formed a table "Mia," and let us call this set of statements "K_{Mia}." Now there is surely a "maximal," coherent set of statements K like K_{Mia} except that every statement in K_{Mia} concerning Mia (or concerning the table formed from m) is replaced by the corresponding statement concerning Woody, and every statement concerning Woody is replaced by the corresponding statement concerning Mia (or the table that actually would have been formed from m), with whatever further additions and deletions are required by these changes. The world W is simply the way-for-things-to-be determined by K. Indeed, W is just like the possible world W_{Mia} corresponding to K_{Mia} (the maximal scenario that would have obtained if m had been formed into a table and Woody had never been constructed), except for the substitution of certain "components" (nonmaximal scenarios, as it were). Since W is a world according to which Woody originates from m, and by hypothesis Woody cannot thus originate, we have here what so many philosophers have so often repudiated: an impossible world. But what is there to repudiate? World W is just the maximal way-for-things-to-be corresponding to a particular set of statements or (potential) facts, something of the same ontological category or sort as the genuinely possible world W_{Mia}. The key difference between W_{Mia} and W is modal rather than ontological-categorical. The former might have been realized whereas the latter could not have been realized; the former is a way things might have been whereas the latter is a way things could not have been. Both are ways for things to be, and in that sense, ontologically on a par.[11]

---

[11]If anything, W_{Mia} is the more dubious of the two, since it directly involves Mia, which does not actually exist, in place of Woody, which actually exists. But let us not worry about this potentially significant ontological difference here. If the truth be told, my own view is that most of the worlds quantified over by modal semanticists do not actually exist, though they might have existed, or possibly might have existed, or possibly possibly might have existed, etc. I do not see this as a decisive reason

IV.

Given this conception of metaphysically possible worlds as forming a restricted subclass of more things of the same ontological category, one cannot rely on the mere existence or nonexistence of worlds according to which it is the case that such-and-such in order to determine whether such-and-such is possible or impossible. It is metaphysically impossible for Woody to originate from m, yet there are many worlds according to which Woody so originates. On my conception, the notions of metaphysical necessity and possibility are not defined or analyzed in terms of the apparatus of possible worlds. The order of analysis is just the reverse: a possible world is understood to be a total way things might have been (or a maximal scenario that might have obtained, etc.), relying on one's prior understanding of the modal notion of what *might have* been. What is possible and what is impossible according to a world is determined by the world itself. Recall that worlds are *maximal* or *total* ways for things to be, deciding all (or a very comprehensive class of) questions of fact. They are not silent concerning all questions of modal fact, since these too are questions of fact. If p is a nonmodal proposition, then one (partial) way for things to be is for p to be a necessary truth, and another is for not-p to be possible. Among the facts (or statements of fact, etc.) that comprise (or obtain according to) a world are such peculiarly modal facts, facts of the form "It is necessary that such-and-such" or "It is possible that such-and-such." It is a fact of the actual world, for example, that it is necessary that Woody does not originate from m, and this fact is included among the facts that comprise the actual world. Given this conception of what a world is, the relevant notion of relative possibility, or accessibility, is perfectly straightforward. If a definition is wanted, it is this: a world w' is *metaphysically possible relative to* a world w if and only if every fact of w' is a possibility in w (that is every proposition that is true according to w' is possible according to w). Equivalently, w' is metaphysically possible relative to w if and only if every necessary fact of w obtains in w' (that is, every proposition that is necessary ac-

not to quantify over them, as long as one keeps one's ontology straight. Cf. my "Existence," in J. Tomberlin, ed., *Philosophical Perspectives I: Metaphysics* (Atascadero, Calif.: Ridgeview, 1987), pp. 49–108.

18

cording to w is true according to w'). If we assume that one question of fact decided by any maximal scenario (or total way for things to be) is the question of whether a given alternative maximal scenario is a scenario that might have obtained (a way things might have been), and we note that on every consistent maximal scenario it itself is the only maximal scenario that obtains, we may prove that every necessary fact of a consistent maximal scenario w obtains in a given alternative maximal scenario w' if and only if on scenario w, w' is a scenario that might have obtained. (If "maximal" scenarios are sets of such things as purported facts, then such facts as that an alternative maximal scenario is a maximal scenario that might have obtained will be meta-facts, which obtain according to the given set of facts not by being included directly as elements of the set but only implicitly by virtue of the facts that are included in the set.) If we confine our attention to consistent maximal scenarios, we may thus put our "definition" another way: to say that a maximal scenario (or total way for things to be) w' is *metaphysically possible relative to* a consistent maximal scenario w is to say that on scenario w, w' is a scenario that might have obtained (a way things might have been). More simply, a world w' is accessible to a consistent world w if and only if w' is possible in w. Being "accessible to" or "possible relative to" a consistent world is simply being possible *according to* that world, nothing more and nothing less. On this conception, what is possible and what is necessary at a given world is not imposed from above by a mysterious and unanalyzed accessibility relation among worlds; rather, a world's accessibility relations to other worlds is internal to the world, *via* the possibilities at that world.[12]

---

[12]Lewis's complaint that "we look in vain, in . . . many . . . places, for an account of what it means to deny that some world is 'relatively possible' " is unjustified. The definition I propose here of the accessibility relation is the natural one, and as Saul Kripke pointed out to me, it follows precisely the characterization of accessibility that he had offered originally in "Semantical Analysis of Modal Logic I: Normal Modal Propositional Calculi," *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik* 9 (1963), pp. 67–96, at p. 70; and again in "Semantical Considerations on Modal Logic," in L. Linsky, ed., *Reference and Modality* (New York, N.Y.: Oxford University Press, 1971), pp. 63–72, at p. 64. There is no suggestion in these pioneering works that such subsystems as T, B, or $S4$ arise from special restrictions on metaphysical modality; instead accessibility is explained in terms of propositions being (metaphysically) possible in worlds. Kripke has informed me (in discussion and personal correspondence) that

It follows, given this conception, that a proposition is metaphysically necessary according to a consistent world w if and only if it is true in every world metaphysically possible relative to w, and a proposition is metaphysically possible according to a consistent world w if and only if it is true in at least one world metaphysically possible relative to w. These are not definitions of metaphysical necessity and possibility. They are theorems that follow from the definition of relative possibility. One must have a prior understanding of metaphysical modality in order to grasp the notion of it being the case that everything that *must* be so on one scenario is so on another scenario (the notion of one world being *possible relative to* another)—as well as the closely related notion of it being the case on one scenario that another scenario is a scenario that *might have* obtained (the notion of one world being *possible according to* another). The idea that the notion of a possible world comes first, and explains the notion of metaphysical modality, is of a piece with the same mythology that gave us the idea that metaphysical necessity is truth in every world whatsoever, without restriction. The notion of metaphysical modality comes first, and like every notion of modality, it is restricted.

There is one alternative yet to be considered. One may choose to ignore ways things could not have been, confining one's sights always and without exception to ways things actually might have been. One may stipulate that a proposition is necessary with respect to an arbitrary possible world w if and only if it is true in every world accessible to the actual world—never mind worlds accessible to w—and likewise that a proposition is possible with re-

---

he is sympathetic to many of the positions advanced in this paper, having seriously considered whether the conventional presupposition that the basic modal logic is *S5* is justified. He now believes he should have stressed both that his use of an accessibility relation does not make "possible" (as applied to worlds) into a dyadic predicate any more than the natural treatment of baldness in possible-world discourse as a binary relation between individuals and worlds makes "is bald" into a dyadic predicate, and that unless we have *S4*, strictly speaking, many of the worlds are not "possible," but only "possibly possible," and so on. Whereas Kripke shares some of my controversial views concerning the logic of metaphysical modality, he is not fully convinced that *S4* modal logic is invalidated in cases like that involving Woody and m (though he tells me he is nearly convinced). Cf. *Naming and Necessity*, p. 51n. See also *Reference and Essence*, pp. 240–252; and "Modal Paradox: Parts and Counterparts, Points and Counterpoints," especially pp. 89–95.

spect to an arbitrary possible world w if and only if it is true in at least one world accessible to the actual world. One may accordingly declare it impossible that Woody even *might have* originated from m, since one is ignoring possibly possible but impossible worlds like W, worlds that are once removed from the actual world on the scale of accessibility and in which Woody originates from m. One may then ignore accessibility altogether. We have finally zeroed in on *S5* modal logic.

This is the ostrich approach to metaphysical modality. It is not a very happy alternative.[13] The ostrich approach flies in the face of the very meanings of the words "necessary" and "possible." On any standard or conventional sense of "possible" in English, a sentence of the form "It is possible that such-and-such" is true if there is a possible (in the same sense) scenario, a way things might have been, according to which it is the case that such-and-such. Certainly this is so with respect to the metaphysical sense of "possible." Likewise, in English, it is simply incorrect to say "It is necessary that such-and-such" when there is a possible scenario according to which it is not the case that such-and-such. In particular, therefore, as long as there is a possible scenario according to which it is possible for Woody to have originated from m, it is true (in English) to say "It is possible that it is possible that Woody originates from m," and one cannot correctly say (in English) "It is necessary that it is necessary that Woody does not originate from m" (or "It is impossible that Woody *might have* originated from m"). If the possible scenarios (such as W') that verify a possibility claim or falsify a necessity claim draw our attention to inaccessible worlds, then we are obliged to pay attention to those inaccessible worlds. We ignore them to our own detriment, counting what is true false and what is false true.

Surprisingly, the ostrich approach has nevertheless ascended to the status of orthodoxy. It is precisely the approach followed by my critics. The most obvious sign of the ostrich approach is the explicit denial of impossible worlds, but there are a number of additional signs, several of which manifest themselves in the objection presented in Section II above. If one ignores impossible worlds altogether, then ways things might have been are the only

---

[13]In *Reference and Essence* I referred (p. 239) to this philosophical position as "a narrow-minded form of modal ethnocentrism."

ways for things to be that are left. The distinction between the generic notion and the modal notion loses all significance. If one confines one's sights to genuinely possible worlds, disavowing the impossible worlds, then metaphysical modality emerges as the limiting case—the "unrestricted" modality that takes account of "every" world—and *S5* emerges as its proper logic. Metaphysical modality appears unrestricted because the restriction to metaphysically possible worlds is already built into one's practice concerning which worlds to pay attention to and to quantify over. If certain entities are ignored entirely and always, then they are not even seen as things that are ignored. Since there is no possible world in which Woody originates from *m*, and possible worlds are the only worlds taken into consideration, one will insist that it is necessary that it is necessary that Woody does not originate from *m*, and that it is necessary that it is necessary that it is necessary . . . that Woody does not so originate, with as many iterations as one pleases. If some iconoclast comes along and argues that some worlds are inaccessible and that in some of these Woody originates from *m*, those who ignore impossible worlds altogether will be puzzled as to what this philosopher could possibly mean by "inaccessible," and hence by "possible" and "might have." Whatever restricted sort of modality the modal iconoclast means by these terms, it would seem to be based on some completely unexplained restriction among the *possible* worlds, for these are the only worlds that are ever considered. When the modal iconoclast protests that in pleading for inaccessible worlds he is not talking about a special and peculiar sort of possible world but about worlds of a sort entirely ignored by the friends of *S5*, those who ignore these worlds will shrug and dismiss these protests as lacking in substance. For in restricting their quantifications over worlds always to possible worlds, they can hardly help but misconstrue the modal iconoclast's claims concerning worlds in general, misinterpreting them as puzzling claims concerning possible worlds. Since he maintains that there are worlds in which Woody does indeed originate from *m*, the modal iconoclast is seen by those who quantify over only possible worlds as capitulating to anti-essentialism. Any such worlds would have to be possible, no matter what the modal iconoclast may mean by calling them "inaccessible," since no other type of world is ever recognized and quantified over, no matter what anyone says. The situation is not unlike that of a philosopher who

22

tries to persuade a pure set theorist, whose quantifiers range only over sets, of the existence of ur-elements (non-set elements), and who is misunderstood as rejecting Extensionality by postulating a plurality of empty sets.

The practice in modal semantics of ignoring worlds that are not possible according to the actual world leads theorists into understanding something different with the use of our terms "necessary" and "possible" from what they mean in English. Specifically, the ostrich approach misconstrues the simple modal term "necessary" to mean the modally complex concept of *actual* necessity, or necessity according to $W_@$, where $W_@$ is the actual world. Likewise, the ostrich approach misconstrues "possible" to mean *actual* possibility, or possibility according to $W_@$. The simple modal concepts of necessity and possibility *simpliciter*—the real meanings of the simple modal terms "necessary" and "possible"—are not the same as the concepts of actual necessity and actual possibility, necessity and possibility according to the actual world. In exactly the same way, the concept of a philosopher is not the same as that of an actual philosopher. The difference shows up in modal contexts. Whereas it was not necessary for Saul Kripke to have been a philosopher, he actually is a philosopher and hence (in the indexical sense of "actually") it is necessary that he be *actually* a philosopher —since in every possible world, the actual world (indexical sense again) is one in which he is a philosopher. Likewise, whereas it is not necessary that it be necessary that Woody not originate from *m*, it is actually necessary that Woody does not so originate, and hence it is necessary that it be *actually* necessary that Woody does not so originate. In effect, the ostrich approach prevents us from speaking of nested modalities altogether, instructing us to misconstrue iterations of modal operators in our speech as redundant embellishments that make no significant contribution to cognitive information content, as mere stuttering. But ignoring impossible worlds does not make them go away, and reinterpreting someone's words to mean what they do not in fact mean does not make the actual meaning go away. Although Woody's originating from *m* is impossible, the presence of worlds such as *W*, in which Woody originates from *m* (and hence, which are impossible) but which are possible according to some possible worlds, makes *something* true as regards the prospect of Woody's so originating. This *something* is expressed in English by saying that the prospect in question is

"possibly possible." The *S5* theorist's misconstrual of English makes nested modality unseen, but it does not make nested modality vanish. The modal iconoclast may echo the words of his colleague: In these questions of haecceitism and essence, by what right do we ignore worlds that are inaccessible? Accessible or not, they're still worlds. Why don't they count? Ignoring all the possibly possible worlds where such-and-such obnoxious things happen, it is impossible that such things even *might* happen. Yes. Small comfort.

The ostrich approach may offer comfort of sorts, but certainly no illumination. It is not I who ignore inaccessible worlds. I acknowledge them and give them their full due, no more and no less. It is my critic, the friend of *S5*, who ignores them altogether. In pleading for inaccessible worlds, I am not drawing an unexplained distinction among the worlds that my opponents recognize, and proposing to ignore those on one side of the undefined boundary line. I am calling attention to worlds to which my opponents pay no attention (other than to repudiate).

V.

The world W, in which Woody originates from m, is a way things could not have been. Nevertheless, there is a way things might have been, W', in which Woody originates from m' instead of from m^*, and in (according to, relative to, from the point of view of) W', W is a way things might have been, as is the way things actually are. The denial of this is highly counterintuitive.[14] The impossible world W is thus only contingently impossible. No doubt it is an essential property of any way things could not have been that it is a way for things to be. And of course, some impossible worlds (such as a world according to which I am a credit card account) are es-

---

[14]I am ignoring here the complications introduced by indeterminacies and regions of vagueness. These complications complicate, but do not significantly alter, the points I am making. Roughly, the idea is that it may in some cases be neither true nor false according to a world w (owing to vagueness in the notion of metaphysical necessity) whether a certain fact obtaining in w is necessary. This, in turn, would inject some indeterminacy into the accessibility relation, so that some worlds may be neither definitely possible nor definitely impossible relative to others. These complications are discussed in some detail in "Modal Paradox: Parts and Counterparts, Points and Counterpoints."

sentially impossible. But others are not. Similarly, it is only a contingent fact about W' that it is a way things might have been rather than a way things could not have been. For there is some matter m'' that Woody might have originated from in lieu of $m*$, and that differs considerably enough from m' (though overlapping just enough with the actual original matter $m*$ to remain a possibility for Woody's origin) that if Woody had originated from m'', it would then have been impossible for Woody to have originated from m'.[15] Let W''' be a possible world in which Woody originates from m''. From the point of view of W''', W' is impossible. Perhaps the actual world is essentially possible. (That is, it may be that the actual world is possible relative to every world possible relative to it.) Even so, some possible worlds are like W', only contingently possible. Whether a world is possible or not can be a question of contingent fact like any other question of contingent fact.

This sort of consideration uncovers the fallacy in the "proof," presented in Section II, of the characteristic $S5$ principle that any possible truth is necessarily possible. The argument was that if a proposition p is true in some possible world w, then no matter what possible world one considers, from its point of view p is true in at least one possible world, namely w, so that in the metaphysical sense of "possible" (in which one possible world in which p is true is all that is required for p to be possible with respect to any given world), if p is possible it is necessarily possible. This argument is framed with an ambiguous usage of the phrase "possible world," indiscriminately meaning either a way for things to be or a way things might have been. The argument is therefore susceptible to

---

[15]As long as some overlap is required and total replacement prohibited, such matter is always possible. Since m' is a possibility for Woody, there will be some overlap between $m*$ and m'. Simply replace as much of $m*$'s overlap with m' as allowable with completely new matter, while preserving the remainder of $m*$, including the entire portion of $m*$ replaced in m'. The resulting matter is m''. It differs from m' by more than the difference between $m*$ and m', since it fully restores all of $m*$'s matter that was replaced in m'—it duplicates the entire difference between $m*$ and m'—and in addition replaces some of the remaining matter of m' with new matter. If the proportion of required overlap is more than one-half (as seems reasonable), some overlap between m'' and m' will remain, but not enough. Since the matter in $m*$ that was replaced in m' has been restored in full, and the maximal replacement by new matter is effected entirely elsewhere in m', the resulting matter m'' exceeds the allowable nonoverlap with m' by exactly the restored matter of $m*$.

two conflicting interpretations. Since our concern is with the logic of what might have been, the argument is of considerably greater philosophical significance when it is interpreted as concerning genuinely possible worlds, rather than worlds in general. Under this interpretation the argument fallaciously presupposes that worlds that are possible in the actual here-and-now are also possible even according to alternative possible worlds.[16] This assumption, though perhaps understandable given the common confusion between possible worlds and worlds in general, is intuitively incorrect. The standard "proofs" of the characteristic B and $S4$ principles likewise involve equivocation between the generic and properly modal sense of the phrase "possible world," resulting in fallacious presuppositions concerning the essentiality of the property of being a possible world (B) or that of not being a possible world ($S4$).

Believers in $S5$ as a correct system of reasoning (in propositional logic) about what might have been must claim that it is an essential property of any way things might have been that things might have been that way. Similarly, believers in the weaker $S4$ modal logic (and hence also believers in $S5$) must claim that it is an essential property of any way things could not have been that things could not have been that way. Believers in B modal logic (and hence also believers in $S5$) must claim that it is an essential property of the way things actually are that things might have been that way. These claims are versions of essentialism. They are doctrines to the effect that certain properties (in this case, certain modal properties) of certain sorts of things (possible worlds, impossible worlds, and the actual world, respectively) are properties that these things could not fail to have. More than this, since their claim is that $S5$, $S4$, or B is a correct *logic* of what might have been and of

---

[16]When the purported "proof" of the characteristic $S5$ principle is interpreted instead (less interestingly) as concerning all worlds without exception, whether genuinely possible or not, it commits a similar error. Under this interpretation the argument fallaciously presupposes that worlds that are available in the actual here-and-now as ones in which a given proposition is true remain available as such even according to alternative impossible worlds. Let w be a world in which a given proposition p is true. One cannot correctly conclude that no matter what world one considers, possible or not, w is still one world in which p is true. There are radically impossible worlds according to which p is not true in w or in any other world.

what must be, the essentialism espoused must be held to be not merely metaphysically true but true by the very *logic* of (metaphysical) necessity and possibility. The essentialism must be held to be not the metaphysically substantive sort of essentialism that requires Woody not to originate from m and me not to be a credit card account, but the minimal, vacuous, and trivial sort of essentialism that requires Woody to be such as to originate or not originate from m, that requires me not to be somebody other than myself, that requires Mars not to be such as to contain life and not to contain life at the same time. This does not weaken the import of the essentialist claims. On the contrary, the logical nature of the claims makes them extremely strong versions of essentialism. The claim is not merely that such-and-such worlds are essentially thus-and-so, but that they are essentially thus-and-so *by logic alone*. It is not merely by virtue of the laws of metaphysics that these worlds are supposed to be essentially thus-and-so, but by virtue of the very laws of logic and nothing more. The doctrine that some properties of some things are properties that on logical grounds alone these things could not fail to have is by itself the most trivial type of essentialism—because it is entirely nonspecific. The doctrine that *such-and-such* properties of *so-and-so* things are properties that on logical grounds alone these things could not fail to have is a horse of a different color. The logical essentialism concerning worlds that the friends of $S5$, $S4$, and B are committed to is some seriously committed essentialism. It is essentialism of the most committed type.

In fact, the logical essentialism concerning worlds that the friends of stronger modal logics are committed to seems intuitively false. At the very least, it requires substantial justification. The possible world W' is a way things are not but might have been; it is a way-for-things-to-be that is not realized, but might have been realized. This is just to say that it is a contingent or accidental feature of W' that it is a way things are not rather than the way things are. I have argued that the accidentalness of the property of being realized is extendible to the modal properties of possibly being realized and of not possibly being realized. Certainly it seems to be logically possible—not precluded by the principles of correct reasoning about modality—that a way-for-things-to-be that might have been realized might have been instead a way-for-things-to-be that could not have been realized, and that a way-for-things-to-be

27

that could not have been realized might have been instead a way-for-things-to-be that might have been realized. The friends of *B* modal logic commit themselves to the loaded claim that it is logically true that the property of possibly being realized (or of being a way things might have been) is an essential property of the actual world. The friends of *S4* modal logic commit themselves to the similarly loaded claim that it is logically true that the property of not possibly being realized is always an essential property of those worlds that have it. The friends of *S5* modal logic commit themselves to the double-barreled claim that it is logically true that both the properties of possibly being realized and of not possibly being realized are always essential properties of the worlds that have them. Yet all admit that the property of being realized is merely an accidental property that possible worlds can have or lack. What, then, is the rationale for their extremely strong versions of logical essentialism? Why should the modal properties of possibly being realized and of not possibly being realized be any less contingent or accidental, from the point of view of pure logic, than the non-modal properties of being realized and of not being realized? These alleged logical truths do not *seem* logically true. Indeed, the last two alleged logical truths, I have argued, are false. The first alleged logical truth, even if it is true, and even if it is necessarily true, does not seem logically true. Surely the burden of proof falls on the logical essentialists with respect to modal properties. We have just seen that the standard "proofs" of the characteristic *B*, *S4*, and *S5* axioms are in fact fallacious, since they assume that any possible world is essentially a possible world (or, in the case of *S4*, that anything that is not a possible world is essentially not a possible world). Whereas this may be trivially true in the generic sense of "possible world," it simply begs the question in the modal sense. The reasoning involved in any purported justification of the contentious doctrine of logical essentialism with respect to modal properties cannot make use of such modal logics as *B*, *S4*, or *S5*—any more than induction can be justified to the Humean skeptic by citing inductive evidence. The systems *B*, *S4*, and *S5* for reasoning about what might have been are precisely what are at issue.

We friends of *T* modal logic are committed to the claim that it is logically true that the actual world has the property of possibly being realized, that as a matter of logic alone, the way things are is a way things might have been. Here we have something that is

28

transparently logically true. Quite plainly, anyone who cannot recognize the validity of an inference from an assertion that it must be that such-and-such to the assertion that such-and-such, does not know how to reason correctly about what must be; and anyone who cannot recognize the validity of an inference from an assertion that such-and-such to the assertion that it might have been that such-and-such does not know how to reason correctly about what might have been. Even the characteristic *B* principle, which may well be necessarily true, does not seem logically true. A proper justification for *B* as a system of modal logic, as opposed to a justificaion for *B* as a *metaphysical theory* of modality, would require not merely a defense of the truth of the essentialist doctrine that the actual world is necessarily possible, and not merely a philosophical argument that the doctrine is indeed a necessary truth, but a convincing case that the doctrine is, like the characteristic principle of *T*, required by logic and nothing more. Until such a justification is provided, modal reasoning in accordance with *B* is not to be recommended—except, of course, insofar as one is prepared to accept a commitment to a certain metaphysical theory. Even then, the *B* "axioms" would not be logical axioms, properly so-called, but metaphysical postulates or premises.

If the modal logical systems *B*, *S4*, and *S5* have never been satisfactorily justified, why are they almost universally accepted as correct systems for reasoning about what must be and what might have been? I have already cited several sources of the present confused state of affairs in contemporary philosophical logic. First, there is the generic-modal ambiguity in the phrase "possible world," which has led to the widely accepted myths that the concepts of metaphysical necessity and possibility are defined in terms of, or constructed from, the concept of a possible world and that metaphysical modality is unrestricted modality. Equivocation between these two senses of "possible world" has led to the fallacious "proofs" of the characteristic *B*, *S4*, and *S5* principles. These fallacious arguments very likely owe something also to another source of confusion in contemporary philosophical logic: the widely adopted ostrich approach to modality, with its consequent misconstrual of "necessarily" as meaning actual necessity and "possibly" as meaning actual possibility. In fact, if the indexical sentential operator "actually" is added to the modal resources of a language, with appropriate logical axioms and restrictions governing its use in

modal reasoning, while retaining only the weak modal system T for the underlying logic of "necessarily" and "possibly," exact analogues to the characteristic axioms and rules of T, B, $S4$, and $S5$ emerge as trivial theorems for the special complex modal operators "actually necessarily" and "actually possibly." In this sense, $S5$ (as the logic of "actually necessarily" and "actually possibly") is a subtheory of T plus the modal logic of "actually."[17] Given its misconstrual of "necessarily" and "possibly," the ostrich approach thus inevitably leads to the acceptance of $S5$ as the correct logic for these modal operators.

My claim is this: the sort of consideration raised in Section I above *demonstrates* the invalidity of $S4$ modal reasoning. I am not proposing a rejection of $S4$ in an *ad hoc* manner, as merely an effective measure for avoiding the difficulty, with no further justification beyond the fact that it avoids the difficulty. The difficulty stems from a widely shared modal intuition, to the effect that some

---

[17]The observation made in the last two sentences derived in part from a fruitful discussion in Dubrovnik with Timothy Williamson (who does not fully endorse the views defended in this article). Williamson correctly observed that although infinitely iterated necessity and infinitely iterated possibility are modal operators for which the analogue of $S4$ is derivable as a subtheory using only T as the underlying logic of "necessarily" and "possibly," the analogues of B and $S5$ are not thus derivable, since the infinitely iterated modalities replace ordinary accessibility by its ancestral, which is automatically transitive but which is not logically required to be symmetric if ordinary accessibility is not. Williamson wondered whether, on my view, there is any modal operator that is definable in terms of "necessarily," and for which the analogue of $S5$ is derivable as a subtheory using only T as the underlying logic of "necessarily." The answer I proposed was: "actually necessarily." (See also note 9 above.)

One characteristic axiom schema of the logic of "actually" is ⌜actually $\phi \supset \Box$ actually ϕ⌝. Another is ⌜$\phi \equiv$ actually ϕ⌝. Application of the rule of necessitation must be restricted to subsidiary proofs that do not invoke the latter axiom.

Williamson's observation generates one serious difficulty for a claim that is often made in response to my arguments and which is closely bound to the myth that metaphysical modality is completely unrestricted: that the logic of necessity and possibility *has to* be $S5$ because "necessarily" really means what I am calling "infinitely iterated necessity" and "possibly" really means what I am calling "infinitely iterated possibility." The logic of what I am calling the "infinitely iterated modalities" would seem to be not $S5$ but $S4$. (A more immediate difficulty with the suggested interpretation is its intrinsic implausibility. For example, it rejects the intuition that, necessarily, Woody might have originated from any wood that is only one molecule different from its original wood but could not have originated from entirely different wood, as not merely false but literally inconsistent.)

small variation in the origin of a material artifact is possible whereas complete variation is impossible. Even if one does not share this intuition, however, it should be quite obvious that the modal position of one (such as myself) who canonizes the intuition into metaphysical doctrine is at least coherent. The position cannot be summarily dismissed on logical grounds alone, as one would (rightly) dismiss the position of someone who proposes restricting the inference rule of *modus ponens* or denying the Law of Noncontradiction or rejecting the characteristic principles of *T*. If the modal position in question seemed not only false but incoherent, a proposal to reject *S4* modal logic solely on the basis of the modal intuition in question would indeed be drastic and poorly motivated. But the mere logical possibility, as opposed to the truth, of the modal intuition is beyond all reasonable doubt. Mere logical possibility, as opposed to truth, is what my argument against *S4* requires. The position outlined in Section I yields a *model* or *interpretation* that both respects the intended interpretation of the logical constants, including "necessarily" (see note 2), and invalidates *S4*. Due consideration of this difficulty makes it *intuitively* plain that *S4* modal reasoning involves a fallacy. Every attempt that I am aware of to retain *S4* modal logic in the face of this difficulty is distinctly counterintuitive.[18] The sort of consideration raised in Section I *exposes* a certain modal *fallacy*, that of inferring the iterated necessity claim "It must be that it must be that such-and-such" from the weaker claim "It must be that such-and-such." Elsewhere I have called this "the fallacy of necessity iteration." This fallacy is the very cornerstone of *S4* modal logic.

University of California, Santa Barbara

---

[18]By far the most popular such attempt is the proposal—made or suggested by Roderick Chisholm, Graeme Forbes, Anil Gupta, Saul Kripke, and Robert Stalnaker (to name but a few)—to replace standard modal semantics with some form or other of counterpart-theoretic modal semantics, as championed by David Lewis. (Kripke's suggestion of a counterpart-theoretic treatment for philosophical problems of the sort engendered by Woody vis-à-vis the matter *m* is made more or less in passing, amid an emphatic rejection of counterpart theory for less problematic modal contexts. See note 12 above.) This alternative system of modal semantics allows for the retention of *S5* modal propositional logic, at a considerable cost. For an accounting of the costs involved, see "Modal Paradox: Parts and Counterparts, Points and Counterpoints."

SELECTED BIBLIOGRAPHY

Adams, R. M., "Theories of Actuality," *Noûs* 8 (1974), pp. 211–231; also in Loux, pp. 190–209.

Adams, R. M., "Primitive Thisness and Primitive Identity," *The Journal of Philosophy* 76 (1979), pp. 5–26.

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

Bringsjord, S., "Are There Set Theoretic Possible Worlds?" *Analysis* 45 (1985), p. 64.

Burke, M. B., "Essentialism and the Identity of Indiscernibles," *Philosophy Research Archives* 9 (1983), pp. 223–243.

Carter, W. R., "Salmon on Artifact Origins and Lost Possibilities," *The Philosophical Review* 92 (1983), pp. 223–231.

Chandler, H., "Plantinga and the Contingently Possible," *Analysis* 36 (1976), pp. 106–109.

Chisholm, R., "Identity Through Possible Worlds: Some Questions," *Noûs* 1 (1967), pp. 1–8; also in Loux, pp. 80–87.

Chisholm, R., "Parts as Essential to Their Wholes," *Review of Metaphysics* 26 (1973), pp. 581–603.

Chisholm, R., "Mereological Essentialism: Some Further Considerations," *Review of Metaphysics* 28 (1975), pp. 477–484.

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

Forbes, G., "Origin and Identity," *Philosophical Studies* 37 (1980), pp. 353–362.

Forbes, G., "On the Philosophical Basis of Essentialist Theories," *The Journal of Philosophical Logic* 10 (February 1981), pp. 73–99.

Forbes, G., "Canonical Counterpart Theory," *Analysis* 42 (1982), pp. 33–37.

Forbes, G., "Thisness and Vagueness," *Synthese* 54 (1983), pp. 235–259.

Forbes, G., "Two Solutions to Chisholm's Paradox," *Philosophical Studies* 46 (1984), pp. 171–187.

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

French, P., Uehling, T., and Wettstein, H., eds., *Midwest Studies in Philosophy XI: Studies in Essentialism* (Minneapolis, Minn.: University of Minnesota Press, 1986).

Fumerton, R. A., "Chandler on the Contingently Possible," *Analysis* 38 (1978), pp. 39–40.

Grim, P., "On Sets and Worlds: A Reply to Menzel," *Analysis* 46 (1987), pp. 186–191.

Gupta, A., *The Logic of Common Nouns* (New Haven, Conn.: Yale University Press, 1980).

Hazen, A., "Counterpart Theoretic Semantics for Modal Logic," *The Journal of Philosophy* 76 (1979), pp. 319–338.

Kaplan, D., "Transworld Heir Lines," in Loux, pp. 88–109.

Kaplan, D., "How to Russell a Frege-Church," *The Journal of Philosophy* 72 (1975), pp. 716–729; also in Loux, pp. 210–224.

Kripke, S., "Semantical Analysis of Modal Logic I: Normal Modal Propositional Calculi," *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik* 9 (1963), pp. 67–96.

Kripke, S., "Semantical Considerations on Modal Logic," *Acta Philosophica*

Fennica 16 (1963), pp. 83–94; also in L. Linsky, ed., *Reference and Modality* (New York, N.Y.: Oxford University Press, 1971), pp. 63–72.

Kripke, S., "Identity and Necessity," in M. Munitz, ed., *Identity and Individuation* (New York, N.Y.: New York University Press, 1971), pp. 135–164; also in S. Schwartz, ed., *Naming, Necessity, and Natural Kinds* (Ithaca, N.Y.: Cornell University Press, 1977), pp. 66–101.

Kripke, S., *Naming and Necessity* (Cambridge, Mass.: Harvard University Press and Oxford, England: Basil Blackwell, 1972).

Lewis, D., "Counterpart Theory and Quantified Modal Logic," *The Journal of Philosophy* 65 (1968), pp. 113–126; also in Loux, pp. 110–128; also in Lewis, 1983, with substantive postscripts, pp. 26–46.

Lewis, D., *Counterfactuals* (Cambridge, Mass.: Harvard University Press, 1973).

Lewis, D., *Philosophical Papers I* (New York, N.Y.: Oxford University Press, 1983).

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

Loux, M., ed., *The Possible and the Actual* (Ithaca, N.Y.: Cornell University Press, 1979).

Menzel, C., "On Set Theoretic Possible Worlds," *Analysis* 46 (1986), pp. 68–72.

McGinn, C., "On the Necessity of Origin," *The Journal of Philosophy* (1976), pp. 127–135.

McMichael, A., "A Problem for Actualism About Possible Worlds," *The Philosophical Review* 92 (1983), pp. 49–66.

Naylor, M. B., "A Note on David Lewis's Realism About Possible Worlds," *Analysis* 46 (1986), pp. 28–29.

Noonan, H. W., "The Necessity of Origin," *Mind* 92 (1983), pp. 1–20.

Odegard, D., "On A Priori Contingency," *Analysis* 36 (1976), pp. 201–203.

Over, D., "The Consequences of Direct Reference," *Philosophical Books* 25 (1984), pp. 1–7.

Plantinga, A., "On Mereological Essentialism," *Review of Metaphysics* 28 (1975), pp. 468–476.

Plantinga, A., *The Nature of Necessity* (Oxford, England: Oxford University Press, 1974).

Quine, W. V. O., "Worlds Away," *The Journal of Philosophy* 73 (1976), pp. 859–863.

Salmon, N., "How *Not* to Derive Essentialism from the Theory of Reference," *The Journal of Philosophy* 76 (1979), pp. 703–725.

Salmon, N., *Reference and Essence* (Princeton, N.J.: Princeton University Press and Oxford, England: Basil Blackwell, 1981).

Salmon, N., "Fregean Theory and the Four Worlds Paradox: A Reply to David Over," *Philosophical Books* 25 (1984), pp. 7–11.

Salmon, N., "Impossible Worlds," *Analysis* 44 (1984), pp. 114–117.

Salmon, N., *Frege's Puzzle* (Cambridge, Mass.: Bradford Books/The MIT Press, 1986).

Salmon, N., "Reflexivity," *The Notre Dame Journal of Formal Logic* 27 (1986), pp. 401–429.

Salmon, N., "Modal Paradox: Parts and Counterparts, Points and Counterpoints," in French, Uehling, and Wettstein, eds., 1986, pp. 75–120.

Salmon, N., "Existence," in J. Tomberlin, ed., *Philosophical Perspectives I: Metaphysics* (Atascadero, Calif.: Ridgeview, 1987), pp. 49–108.

Salmon, N., "Review of David Lewis, *On the Plurality of Worlds*," *The Philosophical Review* 97 (1988), pp. 237–244.

Sharvy, R., "Why a Class Can't Change its Members," *Noûs* 2 (1968), pp. 303–314.

Sharvy, R., "Individuation, Essence and Plenitude," *Philosophical Studies* 44 (1983), pp. 61–70.

Stalnaker, R., "Possible Worlds," *Noûs* 10 (1976), pp. 65–75; also in Loux, pp. 225–234.

Stalnaker, R., "Counterparts and Identity," in French, Uehling, and Wettstein, eds., 1986, pp. 121–140.

van Inwagen, P., "Two Concepts of Possible Worlds," in French, Uehling, and Wettstein, eds., 1986, pp. 185–213.

Wilson, N. L., "Substances Without Substrata," *Review of Metaphysics* 12 (1959), pp. 521–539.

34

Department of Philosophy
1879 Hall
Princeton, New Jersey 08544 U.S.A.

February 3, 1987

Professor Nathan Salmon
Department of Philosophy
University of California, Santa Barbara
Santa Barbara, CA 93106

Dear Nathan,

    Enclosed are two papers, including the one in the Linsky volume too.
The treatment of the relation R is on pp. 69-70 (section 2.1) of
"Semantical Analysis" and on p. 64 of "Semantical Considerations". The
same characterization of R is given in both places. As you see, there is
no suggestion that S5 is basic and the weaker systems come from some
restricted conception. R is characterized in terms of truth and
possibility of propositions in worlds. Notice also the discussion of the
reduction axioms on p. 70 of "Semantical Analysis", and in particular of
transitivity and S4.

    One thing I do is, I now think, somewhat misleading. I should have
stressed that the use of R does not make "possible" (as applied to worlds)
into a two-place predicate, any more than, as you say, "is bald" is.
Probably I only noticed this afterwards. Also, I should have stressed that
strictly speaking, many of the worlds are not "possible" but only "possibly
possible", and so on, unless we have S4.

    By the time I gave the seminar I talked to you about I had definitely
thought these points through, having seriously considered whether the
conventional presupposition that the basic modal logic is S5 is justified.

    I am getting closer to thinking that your treatment of the ship is the
correct solution. Certainly it is a very good piece of work. I am sorry
if almost everyone is unable to see its virtues (you don't say quite that
in the paper). As far as I can see, their counterarguments, as presented,
are confused or circular. It was good talking to you. Talk to you about
Russell, etc., some time.

                                    Best,

                                    Saul Kripke

Enc.

9

This Side of Paradox (1993)

In his intriguing book, *Identity and Discrimination*, Timothy Williamson presents a modified version of a philosophical problem about modality sometimes called 'Chisholm's Paradox'.[1] Williamson proffers a solution based on the apparatus developed in the book, a solution that is at odds with an alternative solution to Chisholm's Paradox that I have defended and developed in a series of essays. Williamson argues[2] that his proposed solution is superior to mine, since it is tailored to handle a variety of philosophical difficulties involving identity, including the original version of Chisholm's Paradox, whereas my solution to the latter involves controversial general claims about modality that are altogether irrelevant to his own version of the paradox. Consider, then, a version of Chisholm's Paradox that I have presented in earlier work.[3] It proceeds from the following two modal principles:

(*A*) If a wooden table *x* is the only table originally formed from a hunk of matter *y* according to a certain plan *P*, and *y′* is any distinct (possibly scattered) hunk of matter that very extensively (sufficiently) overlaps *y* and has exactly the same mass, volume, and chemical composition as *y*, then *x* is such that it might have been the only table originally formed according to the same plan *P* from *y′* instead of from *y*.

(*B*) If a wooden table *x* is the only table originally formed from a hunk of matter *y*, and *z* is any hunk of matter that does not very extensively (sufficiently) overlap *y*, then *x* is such that it could not have been the only table originally formed from *z* instead of from *y*.

 Principle (*A*) is a principle of modal tolerance; principle (*B*) is one of modal intolerance, or essentialism.[4] Chisholm's Paradox starts with the exceedingly

 [1] Timothy Williamson, *Identity and Discrimination* (Oxford: Basil Blackwell, 1990), pp. 126–143. A version of the paradox was apparently first noted by Saul Kripke, in *Naming and Necessity* (Cambridge, Mass.: Harvard University Press, 1980), p. 51 n. 18, where it is briefly discussed. Something directly akin to this paradox was also noted and discussed by Roderick Chisholm in 'Parts as Essential to their Wholes,' *Review of Metaphysics*, 26 (1973), pp. 584–586. The paradox is highly reminiscent of Chisholm's paradoxical queries concerning cross-world identity in his seminal 'Identity Through Possible Worlds: Some Questions,' *Noûs* 1 (1967), pp. 1–8.

 [2] *Identity and Discrimination*, p.p. 127, 135, and 142.

 [3] Nathan Salmon, 'Modal Paradox: Parts and Counterparts, Points and Counterpoints,' *Midwest Studies in Philosophy XI: Studies in Essentialism* (Minneapolis: University of Minnesota Press, 1986), pp. 80–81. See the first endnote of that work for further bibliographical references.

 [4] These are the principles labelled '(II)' and '(III)', respectively, in 'Modal Paradox.' See p. 75 of that work for further modal principles more fundamental than these two. (Thanks to Theodore Guleserian for pointing out the need for a more careful formulation than I had originally given.)

plausible assumption that these two modal principles are not true merely as an accidental matter of contingent fact, but are necessary truths. Furthermore, principle (*A*), at least, is such that if it is true at all, then it is necessary that it is necessarily true, and it is necessary that it is necessary that it is necessarily true, and so on ad infinitum. In fact, on the conventionally accepted system *S*5 of modal propositional logic, any proposition is such that if it is necessarily true, then it is necessary that it is necessarily true, and it is necessary that it is necessary that it is necessarily true, and so on. The paradox consists in a modal propositional argument, which I call '(*CP*)'. The argument, which is valid in *S*5, has numerous premise, all of which seem true, and an explicit contradiction as a conclusion. The first premise is the following:

(P_0) *a* is the only table originally formed from hunk of wood h_0 according to such and such a plan.

This is to be true by hypothesis. Let *n* be the total number of molecules in hunk h_0. We consider a sequence of (possibly scattered) hunks of wood $h_0, h_1, \ldots, h_n$, where each successive hunk of wood in the sequence differs from its predecessor by only one molecule, qualitatively identical to the one it replaces, in such a way that the final hunk h_n has not a single molecule in common with table *a*'s original wood h_0. Premiss (P_0) is then joined by *n* premise of the following form, each of which is derived on the basis of the necessitation of principle (*A*), where $0 \leq i < n:$

(P_{i+1}) Necessarily, if *a* is the only table originally formed from hunk h_i according to such and such a plan, then it is possible that *a* is the only table originally formed instead from hunk h_{i+1} according to the same plan.

These premise are followed finally by the premise,

(P_{n+1}) It is impossible for *a* to be the only table originally formed from hunk h_n according to such and such a plan,

which is derived from principle (*B*). The derivation of the contradictory conjunction of (P_{n+1}) together with that which (P_{n+1}) denies from the premise of (*CP*) is, in some sense, the canonical form of Chisholm's Paradox.

The solution I endorse (following Hugh Chandler) is based on a rejection of the *S*4 axiom, and hence also the *S*5 axiom, of classical modal logic.[5] In its absence, the premise of (*CP*) have no philosophically interesting consequences. A very interesting, and enlightening, consequence is generated, however, if each premise (P_{i+1}) is modified by replacing its initial single occurrence of the modal auxiliary 'necessarily' with *i* or more iterated occurrences—a switch that can be justified on the basis of the infinitely iterated necessitation of (*A*). In the absence of *S*4, the modified premise

---

[5] *Cf.* Hugh Chandler, 'Plantinga and the Contingently Possible,' *Analysis*, 36 (1976), pp. 106–109. The *S*4 axiom is, in effect, the claim that if it is possible that it is possible that *p*, then it is possible that *p*. The *S*5 axiom is the claim that if it is possible that it is necessary that *p*, then it is necessary that *p*. The *B* axiom is the claim that if it is possible that it is necessary that *p*, then *p*. The *S*5 axiom entails both the *B* axiom and the *S*4 axiom in the weak modal logic *T*. In 'The Logic of What Might Have Been,' *The Philosophical Review*, 98 (1989), pp. 3–34, I extend the fundamental argument against *S*4 (and *S*5) into a challenge to *B* propositional modal logic as well. (The work includes a lengthy bibliography.)

taken together with the initial premise (P_0) still do not have the consequence that it is possible for *a* to be the only table originally formed from hunk h_n, but a weaker consequence to the effect that the prospect of *a* being formed from h_n—which, according to (P_{n+1}), is impossible—is nevertheless possibly possibly possibly . . . possible.

Williamson objects that we have no good reason to believe that any of the premise of the canonical version (*CP*) yield counter-examples to the *S*4 axiom:

For [the corresponding premise of analogous] temporal paradoxes are not counter-examples to the analogous principle that if it is at some time the case that it is at some time the case that *A* then it is at some time the case that *A*. They involve the failure of some other assumption; it will have a modal analogue; why should we suppose that the latter does not fail, and blame the *S*4 principle instead? Salmon can point to the intuitive plausibility of the other modal assumptions, but he has not shown it to be any greater than the intuitive plausibility of their temporal analogues, at least one of which is false. For what it is worth, the present author's intuitions are equally strong in the two cases. Furthermore, the *S*4 principle is not behind the modal paradox [presented here].[6]

The crucial wrinkle in Williamson's modified version of Chisholm's Paradox is that we do not begin with an actual artifact. This eliminates altogether the initial premise (P_0) of (*CP*). Instead we are asked to identify and distinguish merely possible artifacts that *would have been* constructed from various portions of matter.[7] A particular carpenter, whose job it is to construct a table from a single hunk of wood according to a specified plan, is repeatedly presented with the entire sequence of hunks $h_0, h_1, \ldots, h_n$ in rapid succession, alternating between sequential order and reverse sequential order. He need only pull a lever in order to select one hunk. Intending to choose at random, the carpenter dies suddenly just before making his selection.[8] Following Williamson's notation, let us abbreviate a modal description of the form 'the merely possible table that would have been the only table originally formed from hunk h_i, according to such and such a plan, had the carpenter selected that hunk and completed the job in that fashion' by '$o(h_i)$'. Intuitively, for each of the descriptions '$o(h_0)$', '$o(h_1)$', and so on, there is a unique possible table that the description designates (assuming each of the terms 'h_i' designates a specific hunk of wood, and ignoring any lingering doubts one may harbor about designating the nonexistent). Furthermore, in considering the differences between the would-be construction of a table from any hunk h_i and that from its immediate successor in the sequence, Williamson argues that, intuitively, such cross-world differences are

too slight to amount to the distinctness of their products. The very same [table] would be made in both cases, but out of marginally different material. . . . The underlying intuition feels

---

6 *Identity and Discrimination*, p. 142.

7 Stephen Yablo informs me that John Drennan had presented a similar version of the paradox.

8 Williamson's actual example involves fashioning a pair of semi-circular earrings by cutting along any diameter of a rotating metal disk. I have taken considerable liberties in modifying Williamson's example to make it more like the situation described in (*CP*). The various differences between Williamson's actual example and my modification of it are not differences on which Williamson places any emphasis. I believe that my modifications do not affect the philosophical points that either Williamson or I wish to make.

the same as that which gives plausibility to somewhat different principles such as Salmon's [modal principle (*A*)].[9]

Let (*W*) be the claim that the cross-world differences between the constructions of tables according to the same plan from neighboring hunks of wood are sufficiently slight to ensure the identity of their products. On its basis we obtain *n* equations of the form '$o(h_i) = o(h_{i+1})$' in place of the former premise (P_{i+1}). In place of the former final premise (P_{n+1}) we have '$o(h_0) \neq o(h_n)$'. Together these new premise entail a new contradiction in classical extensional logic, without any special modal axioms.

Williamson explicitly cites principle (*A*), seemingly approvingly, in support of the *n* equation premise. But recall that he also criticizes my solution to (*CP*), which challenges the modal reasoning involved, partly on the ground that analogous temporal paradoxes impugn the conjunction of modal assumptions involved in the premise of the argument. Williamson has confirmed that he accepts principle (*B*), and hence also the final premise (P_{n+1}) of (*CP*), while rejecting (*A*), or at least its necessitation, and hence also the conjunction of premise (P_1)–(P_n) of (*CP*) which are justified on its basis.[10] His solution to Chisholm's Paradox thus involves embracing a fairly intolerant form of mereological essentialism, in many respects similar to (though perhaps not as extreme as) Chisholm's own brand of essentialism.

Williamson likewise ultimately rejects the conjunction of the first *n* premise in his own version of the paradox. Indeed, in light of the extreme plausibility of the final premise (and the logic of identity), it should be clear that not all of the equation premise can be true.[11] The claim made by (*W*) must be mistaken. Williamson

---

[9] *Identity and Discrimination*, p. 129.   [10] In correspondence, January 1992.

[11] David Cowles has pointed out that the infinite necessitation of (*B*) is insufficient by itself to justify Williamson's final premise that $o(h_0) \neq o(h_n)$. It is logically possible (although very likely metaphysically impossible) that while (*B*) is necessary, and necessarily necessary, etc., the amount of variation possible in the original matter of a typical table exceeds one-half of the totality of its molecules. In that case, all of the first *n* premise may be true. Against this logical possibility, there are at least two ways that Williamson's final premise might be justified. One may simply note that the possible table that would have been the only table originally formed from hunk h_0 if *both* hunks h_0 and h_n had been simultaneously formed into two separate tables, both according to such and such a plan, is none other than $o(h_0)$, and likewise that the possible table that would have been the only table originally formed from hunk h_n if both h_0 and h_n had been simultaneously formed into two separate tables is $0(h_n)$. It immediately follows that $o(h_0) \neq o(h_n)$.

Stewart Cohen and David Cowles have pointed out that this argument does not also show that $o(h_0) \neq o(h_{n-1})$—unless $o(h_{n-1}) = o(h_n)$, or alternatively $o(h_0) = o(h_{n-1})$, where h_{n-1} is a hunk of matter just like h_0 except for the replacement of the one molecule common to both h_0 and $o(h_{n-1})$. In lieu of the above argument, one may invoke a suitable generalization of (*B*), such as the infinitely many principles given by the following schema:

(*B_i*) If *x* is a wooden table and *z* is any hunk of matter that does not very extensively overlap any hunk of matter *y* such that it is possible^i that *x* is the only table originally formed from *y*, then *x* is such that necessarily$^{(i+1)}$, it is not the only table originally formed from *z* instead of from *y*.

Here 'possibly^j' is a string of *j* occurrences of 'possibly', and similarly for 'necessarily^j'. (The original (*B*) corresponds to (*B_0*).) We now make the plausible assumption that the amount of variation possible in the original matter of a typical table is less than one-half of the totality of its molecules. (This assumption may even be strengthened to some extent without significant loss of plausibility.) Let w_0 be any of the 'nearest' possible worlds (those most like the actual world) in

utilizes his rich conceptual machinery to explain why that mistaken assumption seemed plausible.[12]

But he seriously overstates the case when he says categorically that *S*4 modal logic is not behind this problem. There is a clear sense in which what I would deem untrue instances of the *S*4 axiom are precisely what give the problem its air of paradox. I will explain.

Notice first a significant difference between (*CP*) and Williamson's version of the paradox. The latter, but not the former, is formulated in terms of the cross-world identity of possible tables, and indeed the elaborate apparatus that Williamson invokes to explain the intuitive appeal of the mistaken assumption (*W*) is explicitly designed for dealing with cases in which genuine identity is supplanted with certain sorts of approximations to identity. The primary question he poses is: 'Which portions of matter would constitute the same artifact?'[13] This is quite different from the questions posed at the beginning of his discussion of the modal and temporal paradoxes: 'How different could things have been, still being those things? How different could *they* have been?'[14] Although Chisholm originally cast his problem as one concerning identity across possible worlds, and although most others who have discussed the same or related problems (such as Kripke) have also posed those problems in terms of cross-world identity, identity is all but irrelevant to Chisholm's Paradox.[15] Certainly it is not a paradox *about* identity. In particular, the validity of (*CP*), unlike that of Williamson's replacement, does not depend in any way on the logic of identity. As I have argued elsewhere, Chisholm's Paradox is also not a sorites paradox, in the usual sense.[16] It is a paradox about modality.

What of the claimed analogy with the temporal paradoxes? Williamson's contention that the intuitive plausibility of the two modal principles involved in (*CP*) is no greater than that of their temporal analogues is incorrect. Williamson himself,

---

which the carpenter randomly selects hunk h_0. By our assumption, h_n does not sufficiently extensively overlap any hunk h_m that sufficiently extensively overlaps h_0. Hence, by the necessitation of (B_1), instantiated to w_0 (and the double necessitation of (B_0), doubly instantiated to worlds possible relative to w_0), there is no world possible relative to any world possible relative to w_0 in which the actual $o(h_0)$ originates instead from h_n. The actual world is clearly possible relative to w_0. Therefore, none of the nearest worlds in which the carpenter randomly selects hunk h_n is one in which the resulting table is $o(h_0)$.

[12] In the correspondence mentioned above in note 10, Williamson offered a similar account of the plausibility of the necessitation of (*A*). I sharply disagree not only with Williamson's rejection of modal tolerance, but also with this positive component of his account. The positive account includes the claim that each of the *n* equation premise of his own version of Chisholm's Paradox is neither determinately true nor determinately false, because all of the singular terms '$o(h_i)$'—and even much more basic terms like 'that table'— 'fail of perfectly determinate reference' (pp. 133–134, 140–141). An alternative view is that each of the equation premise has a determinate truth-value, though it is not known which it has (over and above the knowledge that some or others are false). In the book Williamson dismisses this view as 'scarcely credible' (p. 133). The former view, in fact, strikes the present writer as far less credible than the latter (partly in light of the central argument of the appendix to 'Modal Paradox,' pp. 110–114), though I am deliberately avoiding these issues here. (Williamson says that he is now more sympathetic to the latter view, though he continues to regard the former as a serious candidate.) [13] *Identity and Discrimination*, p. 131.

[14] *Ibid.*, p. 126.   [15] *Cf.* 'Modal Paradox,' p. 93, last paragraph.
[16] 'Modal Paradox,' p. 89.

like many others, accepts principle (*B*). And he should; it is extremely plausible. In fact, it is surely true. Yet situations like that of the Ship of Theseus pose a very powerful intuitive challenge to a straightforward temporal analogue. Specifically, the familiar tale forcefully challenges the claim that the following is true even of a ship that will undergo extensive refurbishment:

If *x* is the only ship constituted (or the only ship originally constituted) by a hunk of matter *y*, and *z* is any hunk of matter that does not very extensively overlap *y*, then *x* is such that it is never the only ship constituted by *z*.

A great many philosophers share the view that temporal change is more tolerant than modal accident in regard to artifacts and organisms. A table or ship could not have originated from entirely different matter, but once it has been constructed, it is claimed, its material constitution could gradually change, as with a living body, into entirely different matter. Of course, some philosophers (and Williamson is evidently one) favor the status quo, by denying that artifacts have the capacity for total material change.[17] They embrace principles of temporal intolerance, like that displayed above, on intuitive grounds. But then such philosophers should, and probably would, automatically reject temporal analogues of the necessitation of (*A*), on the same grounds. Those grounds strike the present author as comparatively strikingly weak. Perhaps it is not altogether implausible that physical-object artifacts cannot undergo total material change. But just as it is an empirical question whether a living body routinely undergoes gradual total material change, we cannot rule it out *a priori* that tables and ships are forever undergoing rapid total refurbishment right under our very noses—perhaps because of the handiwork of very busy elves, or even of natural processes. By contrast, it does not seem implausible that we can rule it out *a priori* that a table that originated from a hunk of wood might have originated instead from entirely different matter. A priori or not, the conjunction of the necessitations of the original (*A*) and (*B*) is part of my own metaphysical doctrine. It is, at least, a coherent position. Its (relevant) temporal analogue is patently incoherent.

A better temporal analogy to the modal paradoxes arises by replacing the modal auxiliary 'necessarily' with a restricted temporal operator like 'at every moment within the interval from the preceding thirty minutes to the subsequent thirty minutes' and 'possibly' by 'at some moment within the interval from the preceding thirty minutes to the subsequent thirty minutes'.[18] One might then accept appropriate counterparts of the necessitations of both (*A*) and (*B*), even as applied to Theseus's ship.[19] At least they are consistent. Here, of course, the analogue of the *S*4 principle clearly fails.

---

[17] This denial seems somewhat more plausible with regard to such things as languages, as with Williamson's Latin/Italian example (pp. 135–141). I find it considerably implausible with regard to living bodies, and altogether implausible with regard to Heraclitus's river.

[18] *Cf.* my 'Fregean Theory and the Four Worlds Paradox,' in *Philosophical Books*, 25 (1984), pp. 9–10. One may replace the word 'minute' by 'year' or even 'century', if doing so will help to make the point.

[19] In order to obtain the intended assumption, one must change the quantifier on '*y′*' in (*A*) to an existential, change the conditional to a conjunction, etc.

I accept the necessitations of both (*A*) and (*B*), and I argue from their joint truth—or merely from their joint coherence—to the invalidity of *S*4 modal logic.[20] The rejection of *S*4 is not supported merely on the grounds that it provides one way around Chisholm's Paradox. Even if there is a persuasive philosophical argument against principles like (*A*) and (*B*)—and I do not know of any—I would still argue that the position defined by the conjunction of the infinitely iterated necessitations of (*A*) and (*B*) is at least a coherent metaphysical position, and that *S*4 modal logic is thereby *seen to be* fallacious. That metaphysical position demonstrates how it is logically possible for something to be possibly possible without being possible. The mere coherence of the position *exposes* the fallacy in *S*4 modal logic—in something like the way that the overlooked possibility of empty general terms exposes the Aristotelian fallacy of inferring 'Some *S* are *P*' from 'All *S* are *P*'.

I have claimed that Williamson's version of the paradox is driven by the same logical fallacy that drives Chisholm's. Although the argument in Williamson's version of the paradox is classically valid in extensional logic, *S*4 modal logic lies in hiding at the very heart of that paradox. The relevance of *S*4 can be illustrated by means of a convenient (though by no means required) assumption. It is plausible that, although no hunk of wood is actually formed into a table by the carpenter, there is exactly one hunk $h_@$ such that if a selection had been made by the carpenter, it would have been of $h_@$. Notice that the fact that the carpenter would have selected 'at random' does not rule this out. Perhaps Williamson could construct the case in such a way as to rule it out (using quantum indeterminacies or some even stranger device) but pretend for the moment that there is a special such hunk of wood.[21] We may take the possible table that would have resulted from the selection of $h_@$ as having a special modal status—not quite actuality, but the next best thing: being nearest to actuality of all the possible tables in question. This allows us, given sufficient flexibility, to reduce Williamson's possible tables to 'the previous case'; i.e., to a case like (*CP*) in which we begin with an actual table.

Suppose we have the necessitation of the following essentialist principle:

(*A*′) If a wooden table *x* is the only table originally formed from a hunk of matter *y* according to a certain plan *P*, and *y*′ is any hunk of matter that very extensively (sufficiently) overlaps *y* and has exactly the same mass, volume, and chemical composition as *y*, then there could not have been a table that is both distinct from *x* and the only table originally formed according to the same plan *P* from *y*′ instead of from *y*.

Notice that this is a significantly strengthened variant of the original principle (*A*) of modal tolerance, asserting under the relevant hypotheses not merely that *x* might have been the table formed from *y*′ according to plan *P*, but that *x* is the only

---

[20] *Cf.* 'The Logic of What Might Have Been.'

[21] Even if it is assumed instead that several distinct hunks are, so to speak, equally nearly-actual hunks of the carpenter's random selection, if they are close enough to each other in molecular composition (and it is plausible that they will be, as Williamson set up his example—see note 9 above), one may still go some considerable distance along the path we are now on. This is so, in fact, even if there are several such clusters of equally nearly-actual hunks of random selection.

possible table of which this is true.[22] Recall that $o(h_@)$ is the actual-but-for-the-grace-of-God table that would have been constructed had the carpenter lived long enough to finish the job. We may then be willing to say that a selection of a hunk of wood that differs only very slightly from $h_@$ (say by no more than a few molecules) would have resulted in this same nearly-actual table, $o(h_@)$, but that a selection of any hunk of wood that differs from $h_@$ by more than the required margin would have resulted in a different possible table. In fact, this follows from the necessitations of (A') and (B) above, taken together with plausible assumptions to the effect that if it would have been the case, if $o(h_@)$ had existed, that $o(h_@)$ would have been the only table originally formed from y' if y' had been formed into a table, then that actually *is* the case even though $o(h_@)$ does not actually exist; and likewise if it would have been the case, if $o(h_@)$ had existed, that $o(h_@)$ could not have been the only table originally formed from z, then that actually *is* the case even though $o(h_@)$ does not actually exist. One cannot consistently say this, of course, about all the possible tables that might have been constructed by means of a selection from the relevant sequence of hunks of wood. This is what I mean by saying that we are exploiting $o(h_@)$'s near-actuality as the next best thing to actuality. We are assuming that, since $o(h_@)$ is the possible table that would have existed, if any of the relevant possible tables had existed, the relevant limitations on $o(h_@)$'s would-be possibilities (its relevant would-be impossibilities) are also limitations on its actual possibilities. (Of course, one need not attempt to justify the above claims about whether $o(h_@)$ would have resulted from selections of various hunks of wood by means of (A') and (B).)

In saying that the selection of any hunk sufficiently overlapping $h_@$ would have resulted in $o(h_@)$ but that other selections would not have resulted in $o(h_@)$, we thereby reject (W)—an assumption which Williamson defends citing the original principle (A) but ultimately rejects. In fact, even if one rejects the facilitating claim that some hunk of wood is distinguished by being the one that would have been selected, the independent assumption that yields the n equation premise is, as I have already said, clearly untrue in any case. Suppose it were built into the case instead that no hunk in the sequence is distinguished by being a selected-but-for-the-grace-of-God hunk, and that each hunk is instead equally nearly-actual—because of quantum indeterminacies, or whatever. It might then be indeterminable which of the n equation premise is true and which false. But one can still rest assured that some of them are false.[23]

---

[22] Principle (A') is a strengthened variant of a sort of combination of principle (A) and principle (I) from 'Modal Paradox,' p. 75. Under the hypotheses of the principle, hunk y' might have been formed into a table according to plan P, since y' is just like hunk y in all relevant respects. Given (A') together with this observation, the original principle (A) follows. To this extent, (A') is a principle of modal tolerance (as well as a principle of intolerance, or essentialism). Strictly speaking, (A') does not cover Williamson's original example involving possible earrings. (See note 9 above.) In that example, possible artifacts formed by selections of different hunks of matter are not formed, in their respective worlds, according to precisely the same plan, as I had meant the term. But we may construe the term 'plan' more liberally here, so that the same 'plan' is realized in any two such worlds.

[23] This much accords to a significant extent with Williamson's current stance with respect to his problem. See notes 10 and 11 above.

This solution to the problem can be made very similar to—in fact, nearly the same as—the treatment I have proposed elsewhere for a variant of (*CP*) in which each of the *n* premise (P_{i+1}) is replaced by:

(P_{i+1}') If it is possible that *a* is the only table originally formed from hunk h_i according to such and such a plan, then it is also possible that *a* is the only table originally formed instead from hunk h_{i+1} according to the same plan.[24]

This is more like a genuine sorites, or 'slippery slope,' paradox. Here the difficulty is not with the reasoning involved in the argument (which is just modus ponens), but with the premise (P_{i+1}'), not all of which can be true. The suspect modal logical axiom *S*4 remains behind this sorites version of Chisholm's Paradox, however. For one relies on *S*4 in justifying the new premise (P_{i+1}') on the basis of the necessitation of principle (*A*)—or alternatively, on the basis of the legitimately derived former premise (P_{i+1}).[25]

Williamson's argument is much more like this slippery slope variant of (*CP*). The original argument essentially involves nested modality. Williamson might have set up his version of Chisholm's Paradox by citing the necessitation of (*A'*) in lieu of (*W*). In a sense, he should have. By setting it up in this way his problem would have involved nested modality, and thus, would have been significantly more like Chisholm's Paradox, in what I take to be its canonical form. If Williamson will permit it, I also take the result of substituting the necessitation of (*A'*) for (*W*) to be the canonical form of what I hereby dub 'Williamson's Paradox'. It is a deeper, subtler, more paradoxical paradox. This is partly because the necessitation of (*A'*) is enormously plausible—considerably more so than (*W*), which we both reject.

---

[24] The resulting argument is (*CP*)' from section 4 of 'Modal Paradox,' pp. 87–89. See also p. 114 n. 3.

[25] Graeme Forbes suggests justifying the premise (P_{i+1}') independently of *S*4 by means of the following modal principle:

(*F*) If *y'* is any (possibly scattered) hunk of matter that very extensively overlaps a distinct hunk of mattery, and *y'* has exactly the same mass, volume, and chemical composition as *y*, then if a wooden table *x* is such that it might have been the only table originally formed from hunk *y* according to a certain plan *P*, then *x* is also such that it might have been the only table originally formed instead from hunk *y'* according to the same plan *P*.

This principle, which comes very close to (*W*), is equally objectionable. Indeed, given the essentialist principle (*B*), Forbes's principle (*F*) is immediately highly suspicious—and for much the same reason as are the typical general principles from which genuine sorites paradoxes proceed. Compare, for example, the general claim that for any height *h*, and for any distinct height *h'* greater than but very close to *h*, if any adult human with height *h* is short then so is any adult human with height *h'*. One immediately worries about the 'borderline cases': heights *h* and *h'* at or near, or in between, the boundary between being short and not being short. Better yet, consider the claim that for any natural number *n*, if *n* straws did not break the camel's back, then neither will *n* + 1 straws. (Remarks to be made in the final paragraph below concerning the relation between (*W*) and the necessitation of (*A'*) apply, mutatis mutandis, to Forbes's principle (*F*) and the necessitation of the original principle (*A*). In particular, the sharp contrast between the very high degree of plausibility of (*A*) and the evident non-truth of (*F*) casts serious doubt on *S*4.)

This is ironic, since Williamson cites the plausibility of a close variant of (A') as part of the intuitive defense of (W), the assumption he ultimately rejects. It is precisely here that $S4$ comes into play. The necessitation of (A') entails the offending assumption—in $S4$ but not in T. One severs the connection between the switched assumptions by rejecting $S4$. I would suggest that the offending assumption (W) derives much of whatever appeal it may enjoy from the intuitive truth of the necessitation of (A'), and from a failure to distinguish between the two—perhaps as a result of implicitly committing what I call 'the fallacy of necessity iteration' or 'the fallacy of possibility deletion'; i.e., reasoning in accordance with $S4$. This is confirmed by Williamson's explicit citation of a close variant of (A') in his defense of the assumption. Rejecting $S4$ paves the way to rejecting the assumption while retaining the necessitation of (A'). And, of course, rejecting $S4$ provides a solution—indeed, I maintain, the correct solution—to what I take to be the canonical form of Williamson's Paradox.

APPENDIX V • *Cross-World Identification and Stipulation*

47. HAECCEITISM, REDUCTIONISM, AND THE PROBLEM OF CROSS-WORLD IDENTIFICATION

A central topic in the philosophy of modality is *the problem of cross-world identification,* i.e., the problem of identifying individuals in different possible worlds. Consider the possibility of Richard Nixon having continued as United States president for the duration of his second term in office. That is, consider a possible world in which this occurs. We may ask: Would the Democrats have regained the presidency, as they did in the actual world? Would they have nominated Jimmy Carter? And so on. But before we can answer, a philosopher interrupts. What determines whether the president in the possible world under discussion is Nixon? How can we know that it is Nixon rather than someone else who resembles Nixon in a variety of important respects, except for having finished out his presidency rather than resigning in disgrace? And furthermore, what does being Nixon consist in for someone in another possible world? In short, what is the *criterion,* or *criteria,* of cross-world identity that settle the question of whether someone in another possible world is Nixon? In a celebrated critique, Kripke has exposed the alleged problem of cross-world identity as a pseudo-problem (1972, pp. 15–20, 42–53, 76–77). He counters that possible worlds are not like independently existing planets with features to be investigated. "'Possible worlds' are *stipulated,* not *discovered* by powerful telescopes," he says. "There is no reason why we cannot stipulate that, in talking about what would have happened to Nixon in a certain counterfactual situation, we are talking about what would have happened to *him*" (p. 44).

Kripke's contention that possible worlds are "stipulated" has been seriously misunderstood.[1] Many philosophers take it as a thesis

---

[1] A dramatic case in point is Allen Hazen (1979), who asserts (pp. 334–335) that when Kripke says that possible worlds are stipulated rather than discovered, what he means, in part, may be explained by saying that a possible world is a combination of a purely quali-

345

about the ontological and/or epistemological status of possible worlds, about how they came into being and how we come to know of them. They see Kripke as a *modal conceptualist,* who believes that possible worlds are somehow created by us with the properties that we assign to them (a position analogous in certain respects to constructivism about mathematical entities). Readers have thought that Kripke holds that we are the masters of metaphysical modality, in the sense that it is entirely for us to decide, by "stipulation," what is metaphysically possible and what is not. These are serious misinterpretations. Kripke's observation that "possible worlds are not discovered but stipulated" is simply his endorsement of a version of the doctrine that David Kaplan calls *haecceitism.* The *haecceity* of an individual x is the property of being identical with x, i.e., the property of being that very individual. The term 'haecceitism' has been used (perhaps I should say it has been usurped) for a variety of doctrines about possible worlds and the haecceities of the individuals existing in them. This may have been encouraged by Kaplan's official definition. He writes:

> [The doctrine that] we can meaningfully ask whether a possible individual that exists in one possible world also exists in another without taking into account the attributes and behavior of the individuals that exist in the one world and making a comparison with the attributes and behavior of the individuals that exist in the other world . . . [the] doctrine that holds that it does make sense to ask—without reference

tatively specified world together with a particular stipulated choice among various similarity correspondences or mappings (which need not be one-to-one) between individuals in other worlds and individuals of the qualitatively specified world. Hazen thinks of the similarity correspondences as schemes that represent an individual in some other world by means of a selected counterpart in the qualitatively given world. Hazen's entire apparatus is decidedly anti-Kripkean. Kripke adamantly insists that possible worlds need not be purely qualitatively specified, and that the very same individuals may exist in different possible worlds rather than being represented in another world by "counterparts" in that world.

to common attributes and behavior—whether *this* is the same individual in another possible world, that individuals can be extended in logical space (i.e., through possible worlds) in much the way we commonly regard them as being extended in physical space and time, and that a common "thisness" may underlie extreme dissimilarity or distinct thisnesses may underlie great resemblance, I call *Haecceitism.* . . .

The opposite view, *Anti-Haecceitism,* holds that for entities of distinct possible worlds there is no notion of trans-world being. They may, of course, be linked by a common concept and distinguished by another concept—as Eisenhower and Nixon are linked across two moments of time by the concept *the president of the United States* and distinguished, at the same pair of moments, by the concept *the most respected member of his party*—but there are, in general, many concepts linking any such pair and many distinguishing them. Each, in his own setting, may be clothed in attributes which cause them to *resemble* one another closely. But there is no metaphysical reality of sameness or difference which underlies the clothes. . . .

Haecceitism holds that we can meaningfully speak of a thing itself—without reference either explicit, implicit, vague, or precise to individuating concepts (other than being *this* thing), defining qualities, essential attributes, or any other of the paraphernalia that enable us to distinguish one thing from another. It may be that each thing has essential attributes with which it is vested at all times and in each possible world in which it exists. But that is an issue posterior to whether things have trans-world being.[2]

There are at least three distinct doctrines here, each labeled 'haecceitism'. David Lewis takes the central point of haecceitism to be that there are purely qualitatively identical possible worlds that are

[2]Kaplan, 1975, pp. 722–723.

nevertheless distinct in virtue of the individuals represented in those worlds and how they are represented as being.[3] This is a doctrine that Kripke explicitly declines to endorse or reject (ibid., p. 18). Despite the usual gloss on Kaplan's explanations, the central doctrine of haecceitism is not concerned primarily with the identification of individuals in distinct possible worlds—although the doctrine does have important consequences concerning cross-world identifications. The central doctrine primarily concerns an issue of *legitimacy*. It concerns the question of whether it is "meaningful" to stipulate the facts about particular individuals in particular possible worlds, including such facts as that the individual with such-and-such properties in a given world *w* is a particular individual *a*, or is not the particular individual *a*, as the case may be. Haecceitism holds that it is perfectly legitimate when introducing a possible world for consideration and discussion, to specify the world explicitly in terms of facts directly concerning particular individuals, designating those individuals directly by name if one chooses to.

An extreme version of the doctrine—*extreme haecceitism*, as I shall call it—combines haecceitism in the preceding sense with a further doctrine: that facts concerning the particular individual *a* are in some relevant sense primitive, not reducible to any more general facts, such as that the individual with such-and-such properties is thus-and-so. Extreme haecceitism holds that it is legitimate to stipulate facts concerning particular individuals in a world, identifying those individuals by name, precisely *because* such facts about a world are held to be *separate* facts that are not fixed by, and cannot be logically inferred from, facts that do not specify which individuals are involved. I shall use the term 'reductionism' for the opposing doctrine that any such facts about a world *w* as that the individual with such-and-such properties is *a*, or is not *a*, if indeed such facts

---

[3]Lewis (1986, p. 221) compares and contrasts his understanding with a budget of various alternative doctrines that have also gone by the same name of 'haecceitism' (pp. 222–227).

348

exist, are reducible to such qualitative facts as that the individual with such-and-such properties in world *w* is the individual with so-and-so properties in world *w'* (where the so-and-so properties are similar, or closely related, to the such-and-such properties).

Unfortunately, it is unclear what it means to say that facts of one kind are *reducible* to facts of another—or using alternative terminologies, that facts of the first kind "consist in," or are "nothing over and above," facts of the second kind, or that facts of the one kind are "grounded in," "derived from," "based upon," "constructed out of," or "constituted by" facts of the other kind. The central idea seems to be that any fact of the first kind is a *logical* or conceptual consequence of facts of the second kind. An example would help enormously here. But there are precious few, if any, uncontroversial examples. One example from the philosophy of language may do. On Frege's philosophy of semantics, the referential (denotative, designative) facts concerning a language are reducible to other sorts of facts—in particular to intensional-semantic facts about what the sense of an expression is together with extra-linguistic facts about what a given sense metaphysically determines. To illustrate, the English noun 'water', in its use as a name for the familiar liquid, semantically expresses a certain concept (or property) *c* as its English sense, perhaps *the colorless, odorless, potable liquid found (with varying amounts of impurities) in lakes, rivers, and streams*.[4] This is a fact in the theory of meaning—a fact concerning the semantics of sense—and not a fact in the theory of reference. The concept *c*, in turn, metaphysically determines the chemical compound H_2O, in the sense that the compound exactly fits *c* and (let us suppose) no other substance does. This fact is completely independent of language. It is a straightforward logical consequence of these two—the meaning fact and the metaphysical fact—that there is some concept or other such that the word 'water' expresses that concept as its English sense

[4] I use the word 'concept' here in the same sense as Alonzo Church, which is significantly distinct from that of Frege's artificial use of the German '*Begriff*'.

and that concept in turn determines H_2O. The latter, according to a Fregean philosophy of semantics, just *is* the fact that 'water' refers in English to H_2O. This fact is thus partly semantic and partly metaphysical in nature.[5] In this sense, the fact that the English noun 'water' refers to H_2O is "nothing over and above" (consists in, is grounded in, is derived from, etc.) the two facts that the English noun 'water' expresses c and that c determines H_2O.[6]

A doctrine more extreme than simple reductionism opposes

---

[5]In my terminology and conceptual apparatus (Salmon, 1993a, pp. 125–133), the fact in question is (according to Frege's theory of it) a fact of *applied* rather than *pure* semantics, since it involves some extra-linguistic metaphysics.

[6]The notion of reducibility involved here will be clarified further below. An alternative notion of reducibility results by replacing the relation of logical consequence with the notion (metaphor?) of part-whole constitution. We may say that a fact f is *mereologically reducible* to a class of facts c if f is composed, without remainder, of the elements of c. Thus a mereologically complex fact is mereologically reducible to its constituent sub-facts. This notion is suggested by a more literal construal of the terminology of one fact being nothing over and above, or consisting in, etc., a plurality of other facts. The notion presupposes a picture of compound facts as complex wholes resulting from an assemblage of other facts. This picture raises baffling questions about the relationship between mereological reducibility and the logical or conceptual notion of reducibility explicated in the text. On Frege's meta-semantical theory, is the fact that the English word 'water' refers to H_2O mereologically reducible to other facts? In particular, does it mereologically reduce to the pair of facts that 'water' expresses c and that c metaphysically determines H_2O? Is it supposed to be obvious that it does? Suppose 'water' had expressed a different concept in English, but one that also determines H_2O. Would the fact that 'water' refers in English to H_2O then be a different fact, consisting of different sub-facts? Let us say that the proposition that such-and-such, if it is true, *corresponds* to the fact that such-and-such. On some theories, this relation of correspondence is simply identity restricted to true propositions. Suppose that a proposition p corresponds to a mereologically reducible fact f, and that propositions $q_1, q_2, q_3, \ldots$ correspond to the sub-facts to which f mereologically reduces. Is p then logically equivalent to the conjunction $(q_1 \ \& \ q_2 \ \& \ q_3 \ \& \ldots)$? Or is p merely a logical consequence of the conjunction? Or might the two even be logically independent?

Lacking answers to these and other questions, I shall rely in the text primarily on the conceptual notion of reducibility that invokes logical consequence rather than the part-whole relation. It may be useful, however, to bear in mind the possibility that a particular author may instead mean the mereological notion, or something else. Where appropriate, one should distinguish between mereological reductionism and conceptual reductionism (the notion explicated in the text).

350

simple haecceitism. *Anti-haecceitism* is the doctrine that in introducing a possible world for consideration and discussion, one may not legitimately specify facts while mentioning the individuals involved by name (or by something similar, such as by a demonstrative uttered while pointing to an actual individual). Instead, one may specify only the general, qualitative sorts of facts to which the facts concerning a particular individual (if there are any such facts) are reducible according to reductionism. Specifying the facts concerning a particular individual *a*, explicitly identifying *a* by name, is regarded as a form of cheating—or rather, it is held to be meaningless. Some anti-haecceitists go so far as to reject the very existence of such facts about a world as that the individual with such-and-such properties is, or is not, the very individual *a*. They hold that one may not legitimately specify such facts in giving a possible world for the simple reason that there are no such facts to be specified. This view might be called 'extreme anti-haecceitism'. Less extreme anti-haecceitists embrace reductionism, holding that while there are facts directly concerning specific individuals, they are reducible to general facts to the effect that the individual with such-and-such properties is, or is not, the individual with so-and-so properties. Extreme haecceitism in contrast to anti-haecceitism (and in sharp contrast to extreme anti-haecceitism), holds that the former facts are *further* facts over and above general facts, not reducible to or constructed out of the latter. Along with the general facts, these separate facts concerning specific individuals are held to be built into the very fabric of the possible worlds themselves.

Little or no notice has been made in the extant literature on haecceitism of the distinction between the moderate and extreme versions of these various doctrines. I have endeavored to make my usage correspond as closely as possible to established usage of the terms 'haecceitism' and 'anti-haecceitism'. That is why I introduce the special terms, 'extreme haecceitism' and 'reductionism', for the opposing doctrines concerning the question of reducibility (which is

less often the primary focus), and a third term, 'extreme anti-haec-ceitism', for what may be the most controversial of the doctrines. Extreme haecceitism and reductionism are the exact denials of one another. Extreme haecceitism, therefore, might also be called 'anti-reductionism'. One may consistently combine haecceitism (*simpliciter*) with reductionism by holding that it is legitimate to introduce a possible world for consideration by stipulating which facts concerning particular individuals obtain in the world even though such facts are reducible to, or nothing over and above, other sorts of facts. As we shall see, it is possible that Kripke takes this position.

The various versions of haecceitism and anti-haecceitism are perhaps best formulated by invoking a concept from the theory of propositions, that of a *singular proposition*. A singular proposition is a proposition in which at least one individual or object that the proposition is about occurs directly as a constituent, and the proposition is about that individual by virtue of directly including it, rather than a concept by which the individual is represented (determined, denoted). In introducing the terminology of 'singular propositions', Kaplan equates haecceitism with the acceptance of singular propositions (ibid., pp. 724–725). More accurately, haecceitism is the doctrine that one may legitimately cite singular propositions in specifying the propositions that are true in a possible world introduced for discussion. Extreme haecceitism is the stronger doctrine that the truth values of any and all manner of singular propositions are among the primitive, brute facts about which propositions are true and which are false in a given possible world. If one conceives of possible worlds as maximal compossible sets of propositions, then haecceitism holds that possible worlds include singular propositions among their elements in addition to nonsingular, or general, propositions, and extreme haecceitism holds that the entire subset of non-singular propositions included in a world to the effect that the F is such-and-such, for particular properties F, logically entails no singular proposition to the effect that x is such-and-such. Reductionism

holds that the subset of singular propositions, assuming one countenances such propositions at all, is fixed by the subset of nonsingular propositions. Anti-haecceitism (*simpliciter*) holds that possible worlds include *only* general propositions to begin with, leaving open the question of the truth values of any singular propositions, and extreme anti-haecceitism denies that there are any singular propositions to be concerned about.

As Kaplan points out, one should strictly speak of haecceitism, anti-haecceitism, and their variants as relativized to a particular kind of entity *K*, as for example, *anti-haecceitism with regard to concrete things, reductionism with regard to social institutions*, etc. Reductionism with regard to political nations, for example, is the often-cited doctrine that facts involving political nations are reducible to other sorts of facts, such as the actions and histories of particular persons. Extreme haecceitism regarding political nations is the denial of this alleged reducibility. Haecceitism with regard to a kind *K* is logically independent of haecceitism with regard to any logically independent kind *K'*. One may consistently combine haecceitism regarding human bodies with anti-haecceitism regarding persons, for example, by holding that it is legitimate to specify which bodies exist in introducing a possible world for consideration but not to specify which persons exist in that world.

The astute reader will have noticed that I have described the various versions and variants of haecceitism and anti-haecceitism without mentioning the alleged problem of cross-world identification, focusing instead on the role of facts concerning specific individuals in presenting a possible world. How does the cross-world identity problem come in? On anti-haecceitism regarding individuals, possible worlds do not include specific individuals themselves. Instead they provide a structure and framework, given purely qualitatively, in which individuals are represented by means of individual concepts. It is not labeled which individual a given individual concept represents. For the anti-haecceitist, then, there is a special

problem about how the individuals thus represented in distinct possible worlds are to be identified with, or distinguished from, one another. If identification is your game, some assembly is required. And all one has to go on are the individual concepts that represent the individuals. One thus needs criteria of cross-world identity. There is no like problem for the haecceitist, since facts concerning specific individuals may be given directly in specifying the possible worlds under discussion. This is what Kripke means when he says that a possible world need not be given purely qualitatively. Haecceitism holds that facts concerning the haecceities—or in more ordinary parlance, the identities—of specific individuals may be taken as given in introducing a possible world for consideration, and extreme haecceitism holds that all facts concerning specific individuals are directly settled by the internal make-up of the possible worlds themselves. Possible worlds come already equipped with identification labels for the individuals that exist in them. No assembly is required, no identity criteria needed.

Kripke's assertion that possible worlds are not discovered but stipulated is a somewhat less felicitous way of stating what I take to be the central doctrine of haecceitism *simpliciter*, or a closely related doctrine. Criteria for cross-world identity are to be replaced by stipulations. In fact, in this respect possible worlds are no different from anything else that might come under discussion. Suppose I say, "Some cities have monuments made of marble," as a prelude to saying something about some or all such cities. It would be silly (at best) for someone to object that while there are indeed marble monuments in *this* city (the city we are in), I must justify my claim that the monuments in the other cities I have in mind are really made of marble—instead of, say, some other material that was fashioned to look the way marble looks around here. I am discussing cities with marble monuments. I do not have to specify the relevant class of cities purely qualitatively and then provide a criterion for intercity identity of material. I simply select the class of cities that I wish to

discuss by specifying that they have monuments made of . . ., well, *marble*. Kripke contrasts possible worlds, which he says are stipulated, with planets, which are discovered. This may have given the wrong impression. Even independently existing planets may be stipulated in the sense that Kripke intends. One astronomer says to another, "There are undoubtedly thousands of planets that, like Earth, have significant amounts of oxygen in their atmospheres. What is the temperature range for such a planet?" Suppose a philosopher who has been eavesdropping interrupts, "Not so fast. How do you know, and what makes it true, that the atmospheric gas on the planet in question is oxygen, rather than some other element that superficially resembles oxygen? After all, you're not on that planet; you're in no position to send up a weather balloon or to conduct other atmospheric experiments. Are you supposing that, say, atomic number provides a criterion for interplanetary identity of elements? If so, why atomic number? Why not some other feature, like that of having its source in the particular portion of ancient post-Big-Bang material from which our Earthbound oxygen was originally formed?" A reaction by the astronomers of eye-rolling annoyance would be completely justified. The astronomer simply *stipulated* that he was discussing planets that have significant amounts of oxygen in their atmospheres. Even if interplanetary identity criteria for elements are readily available, our astronomer is under no obligation to specify the planets he has in mind purely qualitatively and then ensure that they contain significant amounts of oxygen by providing the available criteria. It is in this sense that even planets are "stipulated." When Kripke says that we do not discover but stipulate possible worlds, he is not making a special claim about their peculiar status vis-à-vis possible worlds. Nor is he claiming that we decree what is possible and what is not. Instead what he means is that the question of which class of possible worlds is under discussion (and in particular the question of which individuals exist in those worlds) is like the matter of which class of entities of any sort is under dis-

cussion, whether they be animals, vegetables, minerals, sticks, stones, or even planets. It is a matter that is entirely open to, and may be entirely governed by, the stipulations of the discussants. The possibility of simply stipulating which individuals are involved renders cross-world identity criteria unnecessary.

48. A RESIDUAL PROBLEM OF CROSS-WORLD IDENTIFICATION

Kripke argues that there remains a problem of cross-world identification even for the haecceitist:

> Although the statement that England fought Germany in 1943 perhaps cannot be *reduced* to any statement about individuals, nevertheless in some sense it is not a fact 'over and above' the collection of all facts about persons, and their behavior over history. The sense in which facts about nations are not facts 'over and above' those about persons can be expressed in the observation that a description of the world mentioning all facts about persons but omitting those about nations can be a complete description of the world, from which the facts about nations follow. Similarly, perhaps, facts about material objects are not facts 'over and above' facts about their constituent molecules. We may then ask, given a description of a non-actualized possible situation in terms of people, whether England still exists in that situation, or whether a certain nation (described, say, as the one where Jones lives) which would exist in that situation, is England. Similarly, given certain counterfactual vicissitudes in the history of the molecules of a table, T, one may ask whether T would exist, in that situation, or whether a certain bunch of molecules, which in that situation would constitute a table, constitute the very same table T. In each case, we seek criteria of identity across possible worlds for certain particulars in terms of those for other, more 'basic', particulars (ibid., p. 50).

What exactly is Kripke's distinction between facts of one kind being "reducible" to those of another, on the one hand, and facts of the first kind merely not being facts "over and above" those of the second, on the other hand? And how does this problem of cross-world identification differ from the pseudo-problem whose illegitimacy is exposed by the observation that possible worlds are stipulated?

Three inter-related notions must be distinguished. We first define *modal supervenience*, as follows:

> Properties of kind K *modally supervene* on properties of kind $K' =_{def}$. For any class c of K-properties and for any class c' of K'-properties, if it is metaphysically possible for there to be something whose K-properties are exactly those in c and whose K'-properties are exactly those in c', then it is metaphysically necessary that anything whose K'-properties are exactly those in c' is such that its (his/her) K-properties are exactly those in c.

Thus, to say that K-properties modally supervene on K'-properties is to say that either it is metaphysically necessary that anything that has exactly such-and-such K'-properties also has exactly so-and-so K-properties, or else it is metaphysically impossible for anything to have exactly such-and-such K-properties and also have exactly so-and-so K-properties. Or put another way, which K-properties a thing has is metaphysically necessitated by which K'-properties it has. For example, to say that a person's psychology modally supervenes on his/her brain and its physical states is to say that a complete accounting of the facts concerning a person's brain and its physical states leaves room for only one possible outcome concerning his/her psychology, in the sense that it would be metaphysically impossible for the person's brain to be in exactly those physical states while the person has a different psychology (even one that is only slightly different).

One may define a notion of *conceptual reducibility* by means of a simple adjustment in the above definition of supervenience,

changing the metaphysical modalities to *conceptual* (or properly *logical*) modalities. It may be assumed here that conceptual necessity entails metaphysical necessity but not vice versa. What is conceptually necessary is true in every conceptually possible world, including such worlds as are metaphysically impossible. To say, then, that properties of kind K are *conceptually reducible to* properties of kind K' is to say that for any class c of K-properties and for any class c' of K'-properties, if it is conceptually (or logically) possible for there to be something whose K-properties are exactly those in c and whose K'-properties are exactly those in c', then it is conceptually (logically) necessary that anything whose K'-properties are exactly those in c' is such that its (his/her) K-properties are exactly those in c. The idea here is that either it is conceptually necessary (a logical or analytic truth) that anything that has exactly such-and-such K'-properties also has exactly so-and-so K-properties or else it is conceptually incoherent (logically inconsistent) for anything to have exactly such-and-such K'-properties and also have exactly so-and-so K-properties. Or put another way, which K-properties a thing has is a logical consequence of which K'-properties it has. For example, on Frege's meta-semantical theory, the referential semantics for a language is reducible to the language's intensional semantics (i.e., its semantics of sense) together with some metaphysics, in that the referential properties of a language are reducible to the language's sense properties taken together with the extra-linguistic matter of what objects are determined by those senses. Given that conceptual necessity entails metaphysical necessity but not vice versa, it follows that conceptual reducibility entails modal supervenience but not vice versa.[7] A claim to the effect that K-properties supervene on K'-properties therefore normally carries the implicature that K-properties are *not* reducible to K'-properties. And indeed, when philosophers explicitly advocate a supervenience thesis, they

[7]Given a certain kind of mereological essentialism, it follows that mereological reducibility of the sort described in the preceding note likewise entails modal supervenience but not vice versa.

often explicitly contrast that thesis with the corresponding reducibility thesis, which they reject, or at least decline to endorse.

This kind of conceptual reducibility must be distinguished from a stronger relative. Let us say that properties of kind K are *strongly conceptually reducible* to properties of kind K' iff for any class c of K-properties there is a class c' of K'-properties such that, by conceptual (logical) necessity, a thing's K'-properties are exactly those in c' if that thing's (his/her) K-properties are exactly those in c. The idea here is that there are particular K'-properties such that it is conceptually necessary (a logical or analytic truth) both that anything that has exactly those K'-properties also has exactly so-and-so K-properties and vice versa. Or put another way, there are K'-properties such that a thing's having those properties is equivalent (logically or analytically) to its having such-and-such K-properties.[8]

Perhaps the most natural interpretation of Kripke's remarks is the following. A table's haecceity—the issue of which table it is, and in particular whether it is the particular table T—is not strongly conceptually reducible to the history of its molecular constitution and the configuration of those molecules—any more than England's haecceity, or its political history, is strongly conceptually reducible to features of the individuals who have made up its population and to their inter-relations with each other and with others. But a table's haecceity is conceptually reducible to its original molecular constitution in the weaker sense. This is the sense in which the table's haecceity is not a fact "over and above" facts about the history of its material constitution. And this weaker kind of conceptual reducibility yields a genuine question concerning cross-world identification of an actual table T. We consider a world—let us call it 's'—in which, it is stipulated, some table or other is the only one ever constructed by configuring exactly such-and-such molecules exactly thus-and-so. We then consider the legitimate issue of whether the table so constructed in s is the actual table T. We are simply asking whether the descrip-

---

[8]Does this notion have the desired consequence that K is not more fundamental than K'?

tion of the table in *s* in terms of its original material constitution entails that the table in question is, or is not, the very table *T*. The conceptual reducibility of a table's identity to its original material constitution ensures that, in some cases at least, there will be an answer to this question. This problem of cross-world identification is different from the pseudo-problem about Nixon. Kripke seems to endorse this new cross-world identification problem as genuine.

There is a problem with this problem. Genuine, full-fledged possible worlds are fully specific with respect to all questions of fact, down to the finest of details. The so-called world *s* is not fully specific in the require way. There are numerous alternative conceptions of what a possible world is. (Not all of these need be thought of as competing conceptions.) The conception I favor is that of a maximally specific scenario that might have obtained.[9] On this conception (and on suitably closely related conceptions), the scenario *s* is the intersection of an infinite plurality of possible worlds, i.e., a constituent "mini-world," or sub-scenario, common to each. It may be regarded as representing the class of those worlds in which some table or other is the only one to originate with exactly such-and-such molecules configured exactly thus-and-so.

Can we simply *stipulate* that the table so formed is *T*? Haecceitism regarding artifacts implies an affirmative answer. More importantly, on extreme haecceitism regarding artifacts, the matter of whether the table so formed is *T should* be stipulated, since the identity (haecceity) of the table is a further fact, not reducible to facts about its original material constitution. If we can simply stipulate that the table so formed is *T*, then we should be equally free to stipulate instead that the table so formed is some table other than *T*. Again, haecceitism regarding artifacts implies that this is indeed so. Of course, the table cannot be both *T* and some table other than *T*. But we are not considering making incompatible stipulations concerning the table's identity simultaneously. We are considering selecting one of them. And why not?

There is no particular reason why not. We can legitimately do this.

[9]Cf. Salmon, 1989a.

As we have seen, *s* represents a class of worlds. That class, it turns out, is diverse. The new problem of cross-world identification—the question concerning the table's identity in *s*—presupposes that in each of the worlds represented by that scenario, the identification goes the same way. This presupposition is erroneous. In some of the worlds represented by *s*, the table so constituted is *T*. In others of those worlds, the table is not. It is illegitimate to ask whether the table in *s* is *T*. This is a matter to be settled by a stipulation concerning which worlds of type *s* are under discussion. We may say, "Consider a world of type *s* in which the table so constituted is *T*." We may also say, "Consider another world of type *s*, different from the last one, in which the table so constituted is this other table, *T'*." Given extreme haecceitism, both sorts of worlds—both of these scenarios—are equally legitimate. They are equally legitimate *qua* scenarios. Neither is incoherent.

The new alleged problem of cross-world identification does not presuppose the controversial thesis that the haecceities of artifacts like tables are strongly conceptually reducible to facts about matter. But it is still every bit a reductionist problem of cross-world identification, since it presupposes the less controversial, but still contentious, thesis that an artifact's haecceity is conceptually reducible in the weaker sense. The alleged problem will be dismissed by the extreme haecceitist (such as myself), and in nearly the same way that Kripke dismissed the more traditional problem—as a pseudo-problem that presupposes a false and unwarranted philosophical doctrine. What I believe to be the correct response to the question raised goes something like this: You, the poser of the question, must tell us which table is the one so constituted. Until you do, you have not provided a genuine world that is specified fully enough to settle the question. It is not for us to *determine* which table is in question. It is up to you to *stipulate* which class of worlds you have in mind. As stated, your question presupposes that the identification of the table so constructed automatically goes the same way for all worlds of type *s*. Since the identification you seek is not reducible to the facts you have given us,

that presupposition is false. Until you make the necessary stipulations, your question is unanswerable in principle. And once you make the necessary stipulations, the answer is then trivial.[10]

Extreme haecceitism (anti-reductionism) makes this kind of dismissal an entirely appropriate response. And in fact, a version of extreme haecceitism is susceptible of something like a proof. Suppose, for a *reductio*, that there is an object x from a possible world w and an object y from a possible world w' such that the fact that $x = y$ is reducible in the weaker sense to (or consists in, is nothing over and above, is derived from, etc.) general facts about x in w and y in w'. (Their identity might be reducible, for example, to x's bearing the relation R in w to a cross-world entity of a certain sort to which y bears R' in w', for appropriate intra-world relations R and R'.[11]) It is evident, by contrast, that the fact that $x = x$ is not similarly reducible to general facts about x in w or in w'. For the fact that $x = x$ is a fact of logic. If it is grounded in any other fact at all, it is grounded only in x's existence (in w or in w'). But then x differs from y in at least one respect. For x lacks y's feature that its identity with x is grounded in general (cross-world) facts about x and it. Conversely, y lacks x's feature that its identity with x is a primitive fact, not grounded in any general facts about x other than its existence. Either way, it follows by Leibniz's Law that x and y are different objects, contradicting the hypothesis that they are identical.[12]

49. A Third Problem of Cross-World Identification

There is a remaining problem of cross-world identification that may arise even for the extreme haecceitist. Imagine that the table T is

[10]Cf. this volume, pp. 242–243.

[11]Cf. this volume, pp. 116–133, on cross-world relations.

[12]See Salmon, 1987a, pp. 517–518. For a variety of controversial, but similarly proved philosophical theses concerning identity, see the appendix to Salmon, 1986c, pp. 110–114 (reprinted in this volume as Section 46 in Appendix IV, p. 325). (Cf. especially T6 and T7 listed there.)

actually fully dismantled and a table is fashioned by configuring exactly such-and-such molecules exactly thus-and-so, as given in the description of *s*. Imagine this really happening. Which table is so constructed? More specifically, is it *T*, or is it some other table?

This is not in any way a matter to be settled by stipulation. Surely there already is some fact of the matter concerning the resulting table's identity with, or distinctness from, *T*. For suppose instead that there is no such fact. Then the table so formed is to that extent unlike *T* (for which there is indeed such a fact). Since they are thus not exactly alike, the table so constituted is distinct from *T*. But then there is a fact of the matter after all.[13] It is not subject to our control what that fact is. If the table so constituted is *T*, that is not at all a

[13]I urged a version of the proof just given in this volume, pp. 242–246. (See also the preceding note 12.) Philosophers who embrace, or otherwise defend, the logical possibility of indeterminate identity have gone to extreme lengths to ward off the counter-proof. Typically, they have responded by accepting that the objects in question differ from each other in the respect cited while rejecting the Leibniz's Law inference from '*a* and *b* are not exactly alike' to '*a* and *b* are not the same thing', on the ground that the conclusion may lack truth value even when the premise is true. The response, however, requires a fundamentally counter-intuitive departure from classical reasoning. For it should be agreed that, of necessity, any *one* thing has every property it has, without exception. It follows by classical reasoning that if the table constituted by such-and-such molecules configured thus-and-so lacks some property that *T* has, then they cannot be *one* table. But if they are not one table, then they are two. (They are certainly not one and one-half tables, for example. Cf. Salmon, 1997.)

[*Homework exercise:* Formalize and derive the preceding argument. What inference rules and/or logical axioms are involved in the derivation? Notice also my use of the plural form '*objects* in question' and of the phrase 'differ from each *other*' in stating the typical response to the original proof. Is this usage consistent with the position stated thereby? If not, is there a coherent way to state the position, in its full generality?]

Derek Parfit (1986, pp. 240–241) endorses the standard reply and says furthermore that even if the proof that there is always a fact of the matter is correct, it only shows that in those cases in which there is no fact of the matter, it is incumbent upon us, if we wish to avoid incoherence, to create a fact by making a decision about the case at hand. This betrays a serious misunderstanding of the proof—and indeed, I believe, a fundamental confusion concerning such things as facts, decisions, and incoherence. The proof demonstrates that there is *already* a fact of the matter, quite independently of any decisions one may wish to make. In addition, a slight variation of the argument shows that it is quite impossible to make a pair of things identical (or distinct) by decision.

result of my (or of our) stipulating that this should be so. No one has made any such stipulation, nor would it have the slightest effect on things if one did. Instead the table's being none other than T seems to be somehow a result of the way the table was constructed, somehow a result of the fact that the table was put together in just this way. The whole business of identity criteria being replaced by Kripkean stipulations seems beside the point, if not completely wide of the mark.

One may feel uneasy about the idea of going beyond mere consideration of the possibility of a given situation, and instead imagining it to be actual. We know it is not actual. Why pretend that it is?

For a simple reason. The point is to mobilize intuitions concerning what *would* be the case if s *had* occurred. If, counterfactually, exactly such-and-such molecules had been configured exactly thus-and-so, then there would be a resulting fact as to whether the table so constituted was T, and that fact would not be a matter of our stipulating what is so. Kripke's observation that "possible worlds are stipulated," properly understood, is simply a recognition of the fact that in considering certain possibilities, we are free to stipulate which possibilities we have in mind by specifying which individuals are involved in them. As we have already seen, it is not a thesis to the effect that what is possible with respect to those individuals is subject to our decision. Nor is it a thesis to the effect that we decide what *would* be the case under certain counterfactual circumstances. There is already a fact of the matter, independently of us, as to which table the table so formed would be if s had occurred. Suppose that the table in question would be T. If so, this appears to be a direct result of the fact that the table was formed by configuring exactly such-and-such molecules exactly thus-and-so. Insofar as it is true that if s had occurred, the resulting table would be T, something significantly stronger is equally true. It is not as if the scenario s might have had different results. If the resulting table would have been T had s occurred, then it is in fact metaphysically *impossible* for s to

occur with the resulting table being some table other than T. In a word, it is *necessary* that the table formed in s is T.

Earlier I said that the class of worlds represented by s is diverse, that there are possible worlds in which s is realized and the resulting table is T and other worlds in which s is realized and the resulting table is some other. Now I am saying that one of these outcomes is impossible, that there are not different possible worlds in which different tables result. I seem to have contradicted myself.

I have not. In previous work I have defended the idea that in whatever sense it is correct and useful to recognize possible worlds as entities, it is equally correct and useful to acknowledge that there are also *impossible worlds*.[14] At this juncture, I invoke the doctrine. Haecceitism does not entail that it is in some way for us to decide what is, and what is not, metaphysically possible. Even extreme haecceitism does not entail this. Haecceitism simply holds that in introducing a world for consideration and discussion, we are free to stipulate the facts that obtain in the world. Depending on what we stipulate, the world, or worlds, we so introduce may turn out to be impossible rather than possible. This is so even if it was our intent to stipulate a possible world. We decide which individuals exist and what properties they have in the world we wish to consider, but metaphysics decides, under its own authority, whether such a world is possible or impossible. The latter issue is completely out of our hands. There are indeed s-type worlds in which the resulting table is T, and there are indeed other s-type worlds in which the resulting table is some other, any other one likes. This is a consequence of extreme haecceitism. The question of the resulting table's haecceity—the question of which table it is—is not to be found among, and does not reduce to or consist in, the facts that are given in the description of s. There are many different ways for the identification to go. But most of those ways are quite impossible. In all of the gen-

---

[14]See Salmon, 1979b, pp. 723–724n; this volume, Section 28 (especially pp. 238–240) and Appendix IV; Salmon, 1984b, 1989a, and 1993b.

uinely *possible* s-type worlds, the resulting table is the same. This is fixed by law but not by legislation. It is fixed by metaphysical law.

It emerges from this analysis that the new problem of cross-world identification bifurcates into two very different problems, differing over whether reductionism is presupposed. A scenario like *s* is first set out, and the question of the resulting table's identity then posed. If the question is put forward under the presupposition of reductionism, it is assumed that one has been given all the facts that are required for deciding the answer, taking the question as concerning *all* the worlds represented by *s*, possible and impossible. One may restrict one's focus to possible worlds, but there is no need to do so. The same answer will obtain for the impossible worlds as well, or at least for the logically consistent ones. For the reductionist, so-called criteria of identity are reductionist analyses or definitions of what it is for a pair of individuals in different worlds to be identical—or at least analytic sufficient conditions for cross-world identity. The question posed is, in effect, an inquiry whether an analytic sufficient condition for cross-world identity of tables has been satisfied. We may call this *the reductionist problem of cross-world identification.* As an extreme haecceitist, I reject this alleged problem as bogus along with the more traditional problem of cross-world identification.

If the question of the resulting table's identity is put forward without presupposing reductionism, one is then presumably being asked to confine one's attention to genuinely possible worlds. In those *possible* worlds in which *s* is realized, which table results? In particular, if *s were* realized, *would* the resulting table be *T*? This question is perfectly legitimate. The facts of the case are sufficient to zero in on one metaphysically necessary outcome. That is to say, even if the resulting table's identity (haecceity) is not even conceptually reducible in the weaker sense to the sorts of facts that one is given in describing *s*, the resulting table's identity does nevertheless supervene modally on exactly such facts.

On the modal-supervenience interpretation of the new problem of

trans-world identification, it is a demand for a metaphysical principle, or principles, that entail the answer to the question of whether the table originally constituted by such-and-such molecules is T. It is, in effect, a demand for T's *essence,* in the sense of a property such that it is metaphysically necessary that a table has the property if and only if it is the very table T and no other. Or perhaps it is a demand merely for a modally *sufficient* property for T's haecceity, i.e., a property such that necessarily, any table with that property is the very table T and no other. Or at the very least, it is a request for an *essential* property of T, i.e., a property that T has necessarily. The sought-after modal property must be adequate to the task of answering the question of the resulting table's relationship to T, interpreted now as a question about genuinely possible worlds in which s obtains. This is *the essentialist problem of cross-world identification,* to be distinguished from the reductionist problem. The essentialist problem does not presuppose that the sort of fact sought in answer to the identity question is strongly conceptually reducible to, or even weakly reducible to ("not a fact over and above"), facts of some other sort. The problem is perfectly compatible with the extreme haecceitist thesis that identity facts are further facts. Even by the extreme haecceitist's lights, it may be seen as a legitimate, and nontrivial, philosophical problem.

In posing a new problem of cross-world identification, does Kripke mean the reductionist problem or the essentialist problem? The textual evidence is inconclusive. He frames his problem explicitly asserting of facts of one kind (the haecceity of the physical object composed of such-and-such molecules) that they are "not facts over and above," and "follow from a description of," facts of another kind (facts about the component molecules themselves). He also cites the traditional reductionist's stock claim that facts about political nations are in this sense nothing over and above facts about people, as an illustration of what he has in mind. And he explains what he means in denying that the facts of the first kind need be "reducible to" the facts

367

of the second, saying that instead there may be "some 'open texture' in the relationship between them," which precludes any "hard and fast identity criteria." These features of Kripke's discussion suggest that he intends the reductionist problem rejected here, as opposed to the extreme-haecceitist/essentialist problem endorsed here.[15]

---

[15]It is possible that Kripke endorses a mereological reductionism of the sort described in note 6 above and that his problem of trans-world identification presupposes this kind of reductionism rather than conceptual reductionism (in the weaker sense). Although Kripke advocates haecceitism in its moderate form, discussions I have had with him (subsequent to the appearance of *Naming and Necessity*) make me doubtful whether he is prepared to hold, as I do, that haecceities are separate from, or facts over and above, such facts about individuals as their molecular composition (though he may be). Cf. Kripke, 1972a, p. 51n; Salmon, 1989a, p. 20n.

Modal paradox II: essence and coherence

Nathan Salmón[1]

Abstract Paradoxes of nested modality, like Chisholm's paradox, rely on *S4* or something stronger as the propositional logic of metaphysical modality. Sarah-Jane Leslie's objection to the resolution of Chisholm's paradox by means of rejection of *S4* modal logic is investigated. A modal notion of essence congenial to Leslie's objection is clarified. An argument is presented in support of Leslie's crucial but unsupported assertion that, on pain of inconsistency, an object's essence is the same in every possible world (in which that object exists). A fallacy in the argument is exposed. Alternative interpretations of Leslie's objection are provided and are found to involve equivocation between different notions of "essence." A material artifact's modal essence, as distinct from its quiddity essence, could have been different than it is.

Keywords Accessibility model · Essence · Essential property · Quiddity · Sarah-Jane Leslie · Modal logic · Modal paradox

1 A modal paradox

I shall say that an object x has a property P *modally essentially* (and that P is a *modally essential property of x*, and is *modally essential to x*) iff it is metaphysically necessary that x has P. An object x is said to have a property P *modally accidentally* (and P is said to be a *modally accidental property of x*, and *modally accidental to x*)

✉ Nathan Salmón
nsalmon@ucsb.edu

[1] Department of Philosophy, University of California, Santa Barbara, USA

iff both x has P and x does not have P modally essentially (i.e., x has P and it is metaphysically possible that x lacks P).[1]

There is a class of paradoxes (antinomies) that invoke nested metaphysical modality and are modal variations on the ship of Theseus.[2] An oversimplified version of one such paradox, often called 'Chisholm's paradox', may be set out as follows. We consider a tripod, which we name 'Troy', and which is the only tripod originally made in the actual world w_1 of three intrinsically purely qualitatively identical interlocking legs L_1, L_2, and L_3. Let L_4 and L_5 be intrinsically similar legs that are distinct from each other and from each of L_1, L_2, and L_3.[3] We assume a principle of origin-tolerance or flexibility, *Tol*, that any tripod with Troy's plan could have been the only tripod made originally from two of the same original legs with a different qualitatively similar third leg. Let 'h' and 'h'' be variables that range over kits consisting of three legs intrinsically similar to L_1; let 'M' be a dyadic predicate for the relation of x being the tripod made originally from a tripod kit h; and let 'O' be a dyadic predicate for tripod kits that have at least two legs in common. We assume also that the relationship of overlap between tripod kits is modally essential to those kits. Then we have:

$$Tol: \quad \forall x \forall h \forall h'[\mathrm{M}(x,h) \& \mathrm{O}(h,h') \rightarrow \Diamond \mathrm{M}(x,h')].$$

Tol is not the sort of principle that can be true only contingently. If true, it is necessary, necessarily necessary, and so on. Indeed, according to the conventionally

---

[1] This terminology differs from that of Leslie (2011). Leslie writes: "An object's essential properties are conditions on what it is to be that object, and this set of conditions fixes just which possibilities or possible worlds the object exists in, namely just those in which it satisfies those conditions. … An object's accidental properties are those of its properties that it can be found without at some times or at some worlds" (p. 277). Leslie's notions appear to be at least partly modal. See note 11.

It is more common to define a modally essential property of x to be a property such that it is metaphysically necessary that x has it if x exists. I believe this to be an error, likely due at least to some extent to the widely held myth that a thing must exist to have properties. The more common definition has two peculiar consequences. First, by this definition a property P can be a "modally essential" property of x, and x can have P "modally essentially," even if x lacks P. For example, had the Eiffel Tower not been erected, it would not have the property of being a tower, though by this definition it would nevertheless have the property "modally essentially." Second, on the more common definition, existence is a "modally essential" property of every possible thing, whereas existence is in fact a modally essential property of some things, e.g., the number two, but not of others, e.g., the Eiffel Tower. (It should be noted that while being a man if existent may be a modally essential property of Socrates, being a man *simpliciter* is not, since Socrates is not a man and does not even exist any longer.)

Fine (1995) says that a property P of an object x is *essential to* x if x must have P to be what it is (p. 53). This definition conflicts with Fine's effort to make 'essence' a term for quiddity essence, and is in fact better suited to the modal notion.

[2] For an overview of the nested-modality paradoxes see (chronologically listed) Chisholm (1967), Kripke (1972), at pp. 50–51; Chandler (1976), Chisholm (1976), at pp. 89–104, Quine (1976), Salmón (1981b), at pp. 229–252, Forbes (1984), Salmón (1986a), Lewis (1986), at pp. 243–248, Kripke (1987), Salmón (1989), Williamson (1990, pp. 126–143), Salmón (1993), Mackie (2006, pp. 59–69) and Robertson Ishii (2013).

[3] According to contemporary philosophical usage, a *world* is a total way for things (the universe) to be. Some philosophers use the phrase 'possible world' incorrectly for a world (in the present sense). A *possible world* is rather a total way things might have been, in the operative sense of 'might have'. Accordingly, an *impossible world* is a total way things could not have been. The *actual world* is the total way things actually are, and is thereby a uniquely distinguished possible world.

accepted system *S5* of propositional logic of metaphysical modality (and according also to the weaker *S4*), any proposition that is necessary is necessarily necessary—hence necessarily necessarily … necessarily necessary, for every number of iterations.

We also assume a principle of modal essentialism, *Ess*, that no tripod of Troy's plan could have been the tripod originally made from only one of the original legs and with two different legs in place of both of the other original legs, let alone from three different legs entirely. Any such tripod in another possible world is not Troy:

$$Ess: \quad \forall x \forall h \forall h\prime [M(x, h) \& \sim O(h, h\prime) \rightarrow \Box \sim M(x, h\prime)].$$

Let 't' be an individual constant for Troy. Let 'h_1' be an individual constant for the tripod kit consisting of L_1, L_2, and L_3; let 'h_2' be an individual constant for the tripod kit consisting of L_2, L_3, and L_4; let 'h_3' be an individual constant for the tripod kit consisting of L_3, L_4, and L_5. The simplified version of Chisholm's paradox is the following derivation:[4]

1.	$M(t, h_1)$	Initial condition
2.	$O(h_1, h_2)$	Initial condition
3.	$\Box[h_2$ exists & h_3 exists $\rightarrow O(h_2, h_3)]$	Initial condition
4.	$\sim O(h_1, h_3)$	Initial condition
5.	$\Diamond M(t, h_2)$	1, 2, *Tol*, logic
6.	$\Diamond\Diamond M(t, h_3)$	3, 5, $\Box$*Tol*, *T* modal logic
7.	$\sim \Diamond M(t, h_3)$	1, 4, *Ess*, *T* modal logic
8.	$\Diamond M(t, h_3)$	6, *S4* modal logic

2 Resolution

Chisholm's paradox is straightforwardly resolved, following Chandler 1976, by rejecting *S4* as the logic of metaphysical modality. Although the prospect of Troy being made originally from h_3 is metaphysically impossible, had Troy been made originally from h_2 instead of h_1, as it might have been, it would have been possible for Troy to have been made originally instead from h_3. Some impossible prospects are such that they might have been possible. The paradox may be seen as a proof that the logic of metaphysical modality is not *S4*, which declares lines 6 and 7 inconsistent. This verdict of inconsistency is intuitively incorrect.

Given *Ess* and $\Box$*Tol*, Troy's modally essential properties are not preserved between possible worlds. In (i.e., according to) the actual world w_1, Troy's property of not being made originally from h_3 is modally essential to it. There is a possible world w_2 (a world accessible to, i.e., possible according to, the actual world w_1) in

---

[4] The formulation ignores largely irrelevant complications arising from the additional premises that $\forall x \forall h \Box[M(x, h) \rightarrow x$ exists & h exists] and $\forall h \forall h' \Box[O(h, h') \rightarrow h$ exists & h' exists].

which Troy is originally made from h_2. In w_2, the property of not being made originally from h_3 is merely modally accidental to Troy. There is a third world, w_3, which is accessible to w_2 and in which Troy is the tripod originally made from h_3. But in the actual world w_1, w_3 is an impossible world.

I shall call this response to Chisholm's paradox '*AR*', an abbreviation for 'the **a**ccessibility **r**esolution'. *AR* accepts *Tol*, $\ulcorner Tol \rightarrow \Box Tol \urcorner$, *Ess*, and lines 1–7 of the paradoxical derivation, while rejecting the *S4* inference at line 8 as an instance of *the modal fallacy of possibility deletion*.

3 Leslie's objection to the foregoing resolution

Sarah-Jane Leslie ([2011](#)) objects that *AR* itself is inconsistent, and that the paradoxes of nested modality are not genuine and therefore not to be taken seriously. Leslie's remarks are quoted here at length with alterations to adapt the remarks to the present example:

> I do not think that it is ultimately satisfactory, for a reason that has not been noted. Salmon's treatment of the paradox faces a destructive dilemma: either the 'paradoxical' argument stops at the second world, in which case there is no paradox to be explained away by Salmon's appeal to the 'deletion fallacy' or he is committed to the view that an item's essence could have been different than it is, even if we restrict our interpretation of the relevant 'could' to the accessible worlds — i.e. the worlds that are possible simpliciter. The world w2 is accessible from w1 and vice versa; each represents straightforward possibilities for the items that exist in the other. But on Salmon's description of the case Troy in w1 has a different essence from Troy in w2. …
>
> What is not possible — not possible simpliciter, since it conflicts with the very notion of essence — is an object having … an essence that varies from possible world to possible world …. An object's essence is its essence in every possible world; any item with a different essence simply cannot be identical to the original object. …
>
> … Salmon's treatment of the paradoxes implies that *Troy's essence could have been different than it is*. We in w1 build Troy with L_1, L_2, and L_3, and agree that Troy's essence is tolerant in that it could have been made with one part different. What this means [*sic*] is that Troy could [only] have been made with two out of those three parts, plus a new [*sic*] part of the relevant sort as needed …. If we accept Salmon's description of the case, then at w2 — where Troy is made from L_2, L_3, and L_4 — Troy's essence is there such that it could [only] have been made from two out of *those* three parts (plus a new [*sic*] part of the relevant sort as needed). But then Troy has an essence at w2 which is *different* from its essence at w1. Since w2 is accessible from w1, we have it that Troy's essence could have been different than it is.

This is just not consistent with the notion of essence. A thing's essence could not have been different than it is. (pp. 284–285)[5]

The word 'essence' is susceptible to a multitude of interpretations in the current philosophical literature. Leslie uses the word for a "combination of essential properties," adding "for the essentialist, an item's essence determines its conditions for existence" (pp. 279–280; see note 1 above). For present purposes, we adopt the following definitions congenial to Leslie's remarks. Where K is any class of properties, we shall say that an object x *has* K iff x has every element of K. We say that x *has* K *modally essentially* iff x has every element of K modally essentially, and that x *has* K *modally accidentally* iff x has K but not modally essentially (x has at least one element of K only modally accidentally). We call the class of x's modally essential properties *the modal essence of* x. Finally, we say that K is a *modal essence* iff K is the modal essence of something or other. On this usage, every object x has exactly one modal essence, and x has its modal essence modally essentially.

One reason for adopting this nomenclature is that it accords with Leslie's claim that in accepting w_3 as a world accessible to w_2 but inaccessible to w_1, *AR* logically entails that Troy's "essence" varies between w_1 and w_2. The nomenclature is strictly a matter of terminology, not of substance. The present terminology, which appears to coincide nearly enough with Leslie's, facilitates the presentation below, but nothing in the analysis depends crucially on the terminology itself. Indeed, each definiendum may be replaced everywhere it occurs by its definiens with no effect on the content of the analysis.

Leslie makes, or appears to make, a couple of puzzling claims. One is that *AR*'s entailing that Troy has varying essences among possible worlds (i.e., among worlds accessible to w_1) had escaped the notice of previous philosophers (*op. cit.*, p. 284). She also claims that previous philosophers—including Hugh Chandler, Roderick Chisholm, Graeme Forbes, Saul Kripke, David Lewis, Penelope Mackie, and yours truly—have uncritically accepted a purportedly inconsistent hypothesis, *to wit*, that *Tol* is true with respect to both w_1 and w_2 (pp. 286–287). To understand why she makes these claims, one must look more closely at Leslie's conclusions. In the quoted passage she states a disjunctive conclusion (mislabeled 'a destructive dilemma'): Either (1) *Tol* is not true with respect to the possible world w_2, so that there is no paradox for *AR* to resolve; or else (2) in advocating *AR*, I am committed to Troy having differing essences among possible worlds. This misstates Leslie's actual conclusion, which is significantly stronger. She in fact argues for the conjunction: (1); furthermore (2).

---

[5] Here and elsewhere, Leslie does not adequately distinguish *Tol* and *Ess*. She says 'what this means' where she evidently intends 'this entails'. The imagined entailment does not in fact exist, however, as *Tol* does not entail *Ess*. The bracketed insertions of the word 'only' yield Leslie's intended instance of *Ess* in lieu of the corresponding cited instance of *Tol*. Leslie says 'a new part' where she should use 'a third part'. (That third part could be the remaining original part, hence not new.) Other interpretations do not yield a viable argument.

Leslie bases her alternative resolution of Chisholm's paradox on a contentious postulation of a plenitude of objects made from exactly the same matter as Troy.[6] Leslie's preferred account has it that in any possible world in which Troy is originally made from h_2 instead of h_1, there is a physical replica of Troy also made from h_2, hence exactly coincident with Troy, such that the replica, but not Troy itself, could have been originally made from h_3. It was stipulated, however, that in w_2, (whatever else there might be) Troy is the only tripod originally made from h_2. Any replicas in w_2 made from the same matter as Troy are not tripods, and hence all but completely irrelevant to the paradox. An inconsistent set of premises cannot be rendered consistent by supplementing the premises with additional theory. A full resolution ultimately must jettison one (or more) of *Ess*, *Tol*, ⌜*Tol* → □*Tol*⌝, and the *S4* principle that whatever is (metaphysically) necessary is necessarily so. Preferably a plausible explanation for the rejected principle's appeal should also be provided. Importantly, it is not sufficient merely to reject *Ess*, for example, or *Tol*, and to provide a weakened substitute. In contrast to the *S4* principle, each of *Ess*, *Tol*, and ⌜*Tol* → □*Tol*⌝ enjoys *prima facie*, pre-theoretic appeal. Those theses, and their consequences, command default assent in the absence of countervailing considerations. A plausible rationale would need to be provided for rejection of any one of them.

Leslie rejects □*Tol*. (It is unclear whether she also rejects *Tol* itself.) Her rationale is a bold charge: □*Tol* by itself is inconsistent. (See note 5.) Leslie calls the property of having varying essences among possible worlds 'a variable essence'.[7] She writes:

> Clearly *Tol* cannot be necessary and true, for consider a pair of mutually accessible worlds in each of which x exists, but which are such that the difference between x's constitution in the two worlds approaches but does not quite meet the allowable limits [*sic*—Leslie means 'does not exceed the allowable limit'] imposed by the requirement of "sufficient substantial overlap". If *Tol* is necessary it follows that x's essential origins are tolerant in the second world in a way that they are not in the first world. That is, it follows that there are possibilities of variable realization of x's essence in the second world that are not found in the first world. This is just what cannot happen, for this implies that x has a variable essence — an essence that changes from world to world — not just a variably realizable essence. (p. 286)

---

[6] Several others have also urged plenitude-centered resolutions. See for example Kment (2014, pp. 194–197). See Robertson Ishii (2013) and Salmón (2018), at $n15$ for responses to Leslie's plenitude.

[7] More precisely, this is the property expressed by '$\lambda x[\sim \exists K \Box(x$ exists $\rightarrow x$'s modal essence $= K)]$', or more perspicuously by '$\lambda x[\exists w \exists w'(w$ is accessible to w' & x's modal essence in $w \neq x$'s modal essence in $w')]$'. Leslie contrasts the property of "a variable essence," which she deems oxymoronic, with a "variably realizable essence," which she defines as an "essence whose fixed fulfillment conditions can admit of varied realization from world to world" (p. 285). She says "none of these paradoxes arise if we distinguish variably realizable 'intolerant' essences and variable essences" (p. 286). The terminology of 'variably realizable essence' and 'a variable essence' encourages the confusion in question. To my knowledge none of the philosophers Leslie criticizes for "sliding from" the former to the latter uses Leslie's terminology or confuses these properties.

Accordingly, Leslie criticizes previous philosophers' uncritical acceptance that if *Tol* is true with respect to w_1 then it is equally true with respect to w_2: "True, once we indulge in this kind of thinking the familiar paradoxes get underway. But the thinking is itself already paradoxical, indeed it is genuinely inconsistent, for it entails that *Troy has a different variably realizable essence depending on whether we start with w1 or w2*. This, once again, is the incoherent idea of a variable essence" (pp. 286–287). What Leslie thinks her predecessors failed to recognize is not merely that *AR* countenances variable essences, but that □*Tol* alone already attributes variable essences, and that therefore □Tol itself is inconsistent (or inconsistent with *Ess*—see again note 5) and the alleged paradoxes of nested modality are not genuine.

Leslie's objection to *AR* can be encapsulated by the following valid argument:

P1: According to *AR*, Troy has a variable essence.
P2: The prospect of an object having a variable essence is inconsistent with the very notion of essence.

Therefore, *AR* is inconsistent.

4 The crux of Leslie's objection

Leslie supports her premise *P1* by observing that according to *AR*, Troy's essence in w_1 includes not being made originally from h_3 whereas Troy's essence in w_2 excludes this same property. The prospect of a variable modal essence is indeed integral to *AR*. No less crucial to Leslie's objection is her premise *P2*. Yet she provides no rationale for *P2*, and it is far from obvious why she believes it.

It is trivial that *AR* depicts Troy as having a variable modal essence; that is indeed the very point of *AR*. It is also trivial that *AR* is consistent. The combination of *Ess*, □*Tol*, and lines 1–4 of the paradoxical derivation, together with their consequence that Troy has a variable modal essence, has a Kripke *B* model—a reflexive, symmetric, non-transitive accessibility model—that interprets all the non-logical constants as intended ('M', 'O', '*t*', etc.). The metaphysical picture painted by this combination of propositions is not merely coherent. It represents a very plausible theory, which many endorse, of the metaphysical facts about material artifacts. Metaphysical necessity is a special way of being true. The characteristic *T* axiom '□$p \rightarrow p$' is straightforwardly analytic if '□' means metaphysical necessity. The characteristic *S4* axiom '□$p \rightarrow$ □□p' does not enjoy this same status. The mere coherence of *AR* places the burden of proof squarely on supporters of *S4* (or something stronger) as the logic of metaphysical modality. Unreserved assertion of *P2* is no substitute for an argument.[8]

---

[8] Leslie (2011) does not acknowledge that there are non-transitive-accessibility models of "variable" modal essences. Perhaps she tacitly assumes that such models violate the logic of metaphysical modality. Admittedly, there are also non-reflexive-accessibility models, and these are inadmissible in the logic of any alethic modality, since they clash with the analyticity of '□$p \rightarrow p$'. Likewise, there are inadmissible

I submit that Leslie's tacit rationale for *P2* commits the informal fallacy of equivocation, perhaps several times over. If her phrase 'a variable essence' is used in a sense on which *P1* is true, *P1* is then quite obvious; and on that sense *P2* is straightforwardly false, as Kripke *B* and *T* models attest. Contrariwise, if the phrase 'a variable essence' is used in a sense on which *P2* is true, on that sense *P1* is simply false and altogether lacking in intuitive support.

A compelling interpretation of Leslie on *P2* arises out of an observation of Teresa Robertson Ishii's: that Leslie appears to slide between the nested modal notion of *x* modally accidentally modally essentially having a property *P* and the incoherent notion of *x* having *P* both modally accidentally *and* modally essentially. More generally, Leslie appears to slide between a proposition *p* being contingently necessary and *p* being both contingent and necessary. One argument in support of *P2* that is strongly suggested by the passage quoted above is the following purportedly logical deduction:

By the definition of 'modal essence' (by "the very notion of essence"):

(1)   For every object *x*, *x*'s modal essence is such that *x* has it in every possible world (in which *x* exists—see footnote 1).
       Therefore,

(2)   For every object *x*, *x* has the same modal essence in every possible world (in which *x* exists).

       Therefore,

(3)   For every object *x*, *x*'s modal essence is the same in every possible world (in which *x* exists).
       The initial observation (1) is correct. However, there is equivocation at (2), which is both ambiguous and slippery. The two relevant readings are given by the following:

(2′)   For every object *x*, there is a modal essence *E* such that *x* has *E* in every possible world (in which *x* exists).
(2″)   For every object *x*, there is a modal essence *E* such that *E* is *x*'s modal essence in every possible world (in which *x* exists).

Unlike (2′), (2″) entails that for every pair of possible worlds *w* and *w′* (in which *x* exists), *x*'s modal essence in *w* is the same as *x*'s modal essence in *w′*. Whereas (2′) is a consequence of (1), it does not yield (3). Alternatively, (2″) delivers (3), but

---

Footnote 8 continued

classical models of 'Jones is married and Jones is a bachelor'. There is a crucial difference between the two modal cases. As regards metaphysical modality, the axioms of *T* are not only analytic but intuitively so. The notion of a false necessary truth is oxymoronic in the same way as the notion of a married bachelor. By contrast, the characteristic axiom of *S4*, '$\Box p \rightarrow \Box\Box p$', is not intuitively analytic; indeed, Leslie is engaged in an on-going controversy concerning its truth. Furthermore, *AR* is clearly coherent, and it poses a forceful case that the characteristic *S4* axiom is even falsified by actual specific instances. The thesis that all instances of that axiom for metaphysical necessity are analytic, so that non-transitive accessibility models are inadmissible, carries the burden of proof and cannot be legitimately assumed, let alone assumed tacitly. *Cf.* Salmón (1989), pp. 28–31.

it is not a consequence of (1). To interpret (2) as $(2'')$ is in effect to treat the phrase 'x's modal essence' in (1) without justification as a rigid designator. It is analytic that every object has its modal essence modally essentially, but there is no inconsistency in the idea that an object might have had what is actually its modal essence without that being its modal essence. Thus $(2'')$ is no mere analytic consequence of $(2')$—unless *S4* (or something stronger) is illicitly assumed as the background modal logic.

The ambiguity of (2) is one of scope, not lexical. There is an alternative potential basis for Leslie's premise *P2*. On the most straightforward interpretation, Leslie does not use 'essence' univocally to mean *modal essence*. (See note 1.) Let us say that x has a property P *logically essentially* iff x has P in every logically possible (i.e., in every consistent) world, whether metaphysically possible or metaphysically impossible; and let *the logical essence of* x be the class of properties that x has logically essentially. (See Salmón 1989.) The hypothesis that by 'essence' Leslie means *logical essence* rather than *modal essence* would explain her repeated assertion without support, as if none is needed, that the prospect of an object having "a variable essence" is incoherent and "conflicts with the very notion of essence." The idea that an object's logical essence somehow varies among possible worlds is indeed extremely dubious.

If one uses the word 'essence' to mean *logical essence*, then it may be correct to say that the notion of "a variable essence" is inconsistent, meaning thereby that the notion of a variable *logical* essence is inconsistent. But it is then incorrect to say that *AR* has the consequence that an artifact has "a variable essence." An object's logical essence is an extremely meager lot compared to the object's modal essence. Troy's logical essence, which includes properties like being either round or not, also includes Troy's haecceity—its thisness, the property of being Troy—and any properties logically entailed by it (e.g., the property shared by Troy and Woody Allen of being either Troy or Woody Allen, and even the property of being either Troy or made originally from h_3). It does not include Troy's more lionized modally essential properties. According to *AR*, Troy has the property of not being made originally from h_3 in every world that is accessible to w_1. *AR* does not cast this property as one that Troy has in every logically possible world. On the contrary, *AR* explicitly depicts Troy as lacking this property in w_3. Leslie sees *AR* as depicting Troy as having "a variable essence," merely on the ground that it denies that the properties that Troy has in every (Troy-inclusive) world accessible to w_1 are the same as those that Troy has in every (Troy-inclusive) world accessible to w_2. It does not follow from this depiction of Troy that the issue of which properties it has in every (Troy-inclusive) logically possible world is somehow relative to w_1 or w_2. The properties that Troy has in every (Troy-inclusive) logically possible world are the very same according to w_1, w_2, and w_3.

Leslie appears to confuse *modal essence* with *logical essence*, with resulting equivocation in her use of the word 'essence'. She argues for her premise *P1* by observing that according to *AR*, not being made originally from h_3 is modally essential to Troy in w_1 but is modally accidental to Troy in w_2. If by 'essence' she means *modal essence*, then her argument for *P1* is correct but her assertion of *P2* is incorrect. If instead by 'essence' she means *logical essence*, then *P2* is justified but

her argument for *P1* is then fallacious. *AR* does not have the consequence, which is of dubious coherence, that Troy has a variable logical essence. If by '*x*'s essence' Leslie means *that which is both x's modal essence and x's logical essence*, then there is a more radical failure. *AR* entails that Troy does not have a modal-*cum*-logical essence.

The very same issues arise in connection with an alternative interpretation of the word 'essence'. As Teresa Robertson Ishii (2013) points out, although Leslie is objecting to *AR*, which employs the notion of an object *x*'s modally essential properties, she sometimes appears to employ instead, or in addition, an Aristotelian notion of "essence" urged by Joseph Almog, Kit Fine, and Stephen Yablo: *what x is*, or what it is to be *x* (pretending these are the same thing). An object's *quiddity essence*—its whatness—is supposed to be a very select, privileged segment of the object's modally essential properties (but presumably not merely the object's haecceity).[9] If there is such a thing as Troy's quiddity essence—if Troy has a whatness (and only one)—it is a severely restricted subclass of Troy's modal essence. As Robertson Ishii notes, Leslie's terminology blurs together modal essence and quiddity essence. Contrary to some grandiose claims the alternative uses of 'essence' are not competing, but they are different. The modal concept is clear; the quiddity notion is unclear. But it is clear that what has just been said concerning an object's logical essence is equally true of its alleged quiddity essence. The idea that Troy's quiddity essence varies among metaphysically possible worlds is indeed of dubious coherence. Fine (2005), pp. 348–349 writes that

> the identity of an object is independent of how things turn out, … not just in the relatively trivial sense that the identity of an object is something that will hold of necessity. Rather it is the core essential features of the object that will be independent of how things turn out and they will be independent in the sense of holding *regardless* of the circumstances, not *whatever* the circumstances. The objects enter the world with their identity predetermined, as it were; and there is nothing in how things are that can have any bearing on what they are.

I take it that Fine means to say this: The quiddity essence of a possible object is independent of the circumstances of a world, not only in the trivial sense of 'holds in every possible world whatever its circumstances', but furthermore in the stronger sense of 'holds in a possible world quite independently of its circumstances'. A possible object supposedly has its quiddity essence in every possible world precisely *because* quiddity essence is world-independent. The world-invariance is a consequence of the world-independence.[10]

If by 'essence' Leslie means *quiddity essence*, then again, even if *P2* is justifiable her argument for *P1* is fallacious. Kripke models demonstrate that varying modal essences are consistent with fixed quiddity essences. It is perfectly compatible with

---

[9] Yablo (1987), Almog (1991) and Fine (1994, 1995). Fine also refers to the quiddity essence of an object *x* as *x*'s "identity" (a term better suited to *x*'s haecceity) and *x*'s "definition".

[10] It will not do for Fine's purposes to allow that a property that is part of an object's quiddity essence can be a metaphysically contingent property of the object.

AR that although Troy's modal essence varies among w_1, w_2, and w_3, Troy's quiddity essence—what it is to be Troy—supposing Troy has such a thing, is the very same in all three worlds.[11]

In fact, it would seem that Troy's quiddity essence (assuming it has one) must be the same in all transitively metaphysically possible worlds (where a transitively metaphysically possible world is one that is metaphysically possible, metaphysically possibly metaphysically possible, or so on). For if Troy's quiddity—if what it is to be Troy—were different in any pair of worlds w and w', Troy would not be in w' exactly what it is in w, *to wit*, that very *tripod*. In both worlds, Troy has the property of being a tripod, for example, and in both worlds Troy has the very same haecceity, the property of being *this very* thing.

It should be noted that contrary to Fine, objects can nevertheless lose even their quiddity essences in some far away worlds. In a logically possible world in which Troy is a credit-card account or a poem instead of a tripod—a metaphysically impossible way for things to be that is not even transitively possible—Troy presumably lacks the quiddity essence that it has (assuming it has one) in w_1, w_2, and w_3. (Troy retains its haecceity even in such far away worlds.)

5 A final interpretation

It is possible that Leslie uses the phrase 'a variable essence' altogether differently. In an alternative nomenclature one might define '*the essence of*' an object x in a world w to be the class of properties that x *actually* has modally essentially, i.e., the class of properties P such that x has P in every world w' that is accessible to the actual world w_1 (rather than to w). In the terminology of the present essay this is x's *actual modal essence*, i.e., the modal essence of x in the specific world w_1. Of course, the actual modal essence of an object x is just x's modal essence, nothing more and nothing less. For this reason, it is easy to confuse the notion of an object's modal essence with that of an object's actual modal essence. There is a very important difference between the two notions. The difference shows itself in non-actual worlds, e.g., in merely possible worlds like w_2. It is an element of Troy's actual modal essence that Troy not be made originally from h_3. This same property—not being made originally from h_3—is not an element of Troy's modal essence in w_2, but it remains an element of Troy's actual modal essence in every world, including w_2.

---

[11] One candidate for Troy's quiddity essence is the property of being both Troy and a tripod, or the pair set consisting of the property of being a tripod together with Troy's haecceity (being Troy).

Leslie might endorse the highly dubious thesis that an object's quiddity essence determines the object's full modal essence, and it is possible that this thesis is part of Leslie's rationale for *P2*. Numerous philosophers are committed to a weaker thesis: that the collective quiddity essence of (the plurality of) all objects determines the modal essence of each object. As Robertson Ishii observes, *AR* poses a very formidable challenge even to this weaker thesis. If Leslie endorses either thesis, she is not entitled simply to assume it, especially not as a tacit premise. Both the weaker thesis and the stronger thesis bear the burden of proof. *AR* disputes both theses, and more importantly, it presents a very forceful case against even the weaker thesis. (See note 8.)

What is actual (in the indexical sense) is actual in every world; the idea that an object's actual modal essence varies among different worlds is incoherent. That Troy's actual modal essence is invariant among (Troy-inclusive) possible worlds is a potential basis for Leslie's assertion of *P2*. The prospect of an object having a variable actual modal essence is indeed inconsistent with the very notion of an actual modal essence.

This is not to say that Leslie and I mean different things by 'essence' so that the differences between us are merely verbal. Leslie and I sharply disagree on matters of both metaphysical and modal-logical substance. All parties should agree that Troy's entire actual modal essence is the same in w_2 as it is in w_1. In particular, the property of not being made originally from h_3 is as much an element of Troy's actual modal essence in w_2 as in w_1. *AR* also has it that in w_2, Troy could have been made originally from h_3, i.e., if w_2 had been realized—as it might have been—then it would have been possible for Troy to have been made originally from h_3. That Troy's actual modal essence precludes the prospect that Troy is made originally from h_3 merely confirms that Troy's modal essence in w_2 is different from Troy's actual modal essence (i.e., from Troy's modal essence in w_1). By contrast, Leslie contends that insofar as Troy could not have been made originally from h_3, even if Troy had been made originally from h_2, it would still have been metaphysically impossible (by the very notion of "essence") for Troy to be made originally from h_3.

It is a consequence of Leslie's view of the matter that given *Ess*, the putatively metaphysically impossible world w_3—which she stipulated to be a world (assuming there is one) in which Troy is made originally from h_3 (p. 283)—is impossible even according to w_2. Yet Leslie asserts that Troy does not exist in w_3 (pp. 288–289), directly contrary to her stipulation. This provides indirect evidence that by 'essence' she means *actual modal essence*. Leslie mistakes the stipulated impossible world w_3, which includes Troy, for a possible world w_1' in which Troy is absent and a different tripod, Trevor, is made from h_3. In effect, Leslie misidentifies w_3 with its metaphysically possible twin. (On Leslie's view, w_1' is also possible according to w_2. On my view, it is not.) The likely explanation for her confusion of w_3 with w_1' is that Leslie does not recognize actually impossible worlds and sees only actually possible worlds.[12] Whereas w_3 is not among the possible worlds, w_1' is. When her attention is directed toward the impossible world w_3, Leslie attends instead to its possible counterpart.

In (1989) I referred to the general confusion of the notion of necessity with that of actual necessity, as 'the ostrich approach to metaphysical modality', because it fails to acknowledge worlds like w_3 that are possibly possible but not possible *simpliciter*. The ostrich approach maintains the discredited S5 as its modal propositional logic by ignoring worlds inaccessible to the actual world. The "logical

---

[12] The four-world paradox compares w_2, which includes Troy, with a world w_2' that is possible according to w_1' and in which Trevor instead of Troy is made from h_2. David Lewis (1986, p. 245*n*) in effect also misidentifies w_3 with w_1', for the same reason as Leslie but supported also by his highly idiosyncratic understanding of metaphysical modality (which I deem a colossal misunderstanding).

space" of the ostrich approach is metaphysically impoverished. It is missing a plenitude of impossible worlds.

If by 'essence' Leslie means *actual modal essence* (or *modal-essence-cum-actual-modal-essence*), then her assertion of *P2* is justified but her argument for *P1* is fallacious. *AR* does not have the incoherent consequence that Troy has a variable actual modal essence.

It is possible that Leslie does not equivocate with (2) (or anything similar) in support of her crucial premise *P2*. It is possible that she does not confuse *modal essence* with *logical essence* or with *quiddity essence* or with *actual modal essence*. It is equally possible that she equivocates in all these ways. If her tacit rationale for *P2* is not mistaken in any of these ways (nor in the way criticized in note 8), then I am unable to guess what that rationale is.

It is inconceivable that Kripke, Lewis, and others who have addressed *AR* were all unaware that it has the consequence that an artifact might have had some of its actually modally essential properties merely modally accidentally. The reason they had not noticed that *AR* is inconsistent is that it is consistent. The intended Kripke model establishes consistency. (See notes 7 and 8.) Leslie's rationale for rejecting □*Tol* thus collapses. *AR* is not merely consistent in *T* and *B* modal logics. Its core theses—*Tol*, ⌜*Tol* → □*Tol*⌝, and *Ess*—reflect metaphysical common sense. Artifacts, and presumably also material objects of some natural kinds, genuinely have different properties modally essentially in different possible worlds. The modal essence of any material artifact genuinely could have been different than it is.

Leslie says that *AR* runs a gamut, being at once ingenious, influential, based on confusion, incoherent, and inconsistent (pp. 283–287). As far as I am able to determine, *AR* is in fact none of the above. Numerous philosophers who reason with modality persist in embracing *S5* as the presumed propositional logic of metaphysical modality. Allegiance to *S5* modal logic notwithstanding, the axioms and rules of *S5* were not handed down unto us engraved on sacred tablets. That it is at least logically possible that some *de re* metaphysical necessities are only contingently necessary—so that the logic of what might have been is not even as strong as *S4*—is little more, but nothing less, than good philosophical sense.

References

Almog, J. (1991). The what and the how. *Journal of Philosophy, 88*(5), 225–244.

Chandler, H. S. (1976). Plantinga and the contingently possible. *Analysis, 36,* 106–109.

Chisholm, R. M. (1967). Identity through possible worlds: Some questions. *Noûs, 1,* 1–8.

Chisholm, R. M. (1976). *Person and object: A metaphysical study*. La Salle: Open Court.

Fine, K. (1994). Essence and modality. In J. Tomberlin (Ed.), *Philosophical perspectives. Logic and language* (Vol. 8, pp. 1–16). Atascadero: Ridgeview Publishing Co.

Fine, K. (1995). Senses of essence. In W. Sinnott-Armstrong, D. Raffman, & N. Asher (Eds.), *Modality, morality, and belief* (pp. 53–73). Cambridge: Cambridge University Press.

Fine, K. (2005). Necessity and non-existence. In K. Fine (Ed.), *Modality and tense* (pp. 321–354). Oxford: Oxford University Press.

Forbes, G. (1984). Two solutions to Chisholm's paradox. *Philosophical Studies, 46*(2), 171–187.

Kment, B. (2014). *Modality and explanatory reasoning*. Oxford: Oxford University Press.

Kripke, S. A. (1972). *Naming and necessity*. Oxford: Basil Blackwell.

Kripke, S. A. (1987). Letter to Salmón concerning Salmón 1989 (unpublished).

Leslie, S.-J. (2011). Essence, plenitude, and paradox. In J. Hawthorne & J. Turner (Eds.), *Philosophical perspectives, 25, metaphysics* (pp. 277–296). New York: Routledge.

Lewis, D. (1986). *On the plurality of worlds*. Oxford: Basil Blackwell.

Mackie, P. (2006). *How things might have Benn*. Oxford: Oxford University Press.

Quine, W. O. (1976). Worlds away. *Journal of Philosophy, 73*(22), 859–863.

Robertson Ishii, T. (2013). Everything but the kitchen sink: How (not) to give a plenitudist solution to CP. In *Presented orally at numerous venues in 2013 and subsequently*.

Salmón, N. (1981a). Modal paradox: Parts and counterparts, points and counterpoints. In P. French, T. Uehling, & H. Wettstein (Eds.), *Midwest studies in philosophy XI: Studies in essentialism* (pp. 75–120). Minneapolis: University of Minnesota Press; reprinted in Salmón 2005*a*, pp. 273–344.

Salmón, N. (1981b). *Reference and essence* (2nd ed.). Amherst, NY: Prometheus Books.

Salmón, N. (1989). The logic of what might have been. *Philosophical Review, 98*(1), 3–34; reprinted in Salmón 2005, chapter 7, pp. 129–149.

Salmón, N. (1993). This side of paradox. *Philosophical Topics, 21*(2), 187–197; reprinted in reprinted in Salmón 2005, chapter 9, pp. 155–164.

Salmón, N. (2005). *Metaphysics, mathematics, and meaning*. Oxford: Oxford University Press.

Salmón, N. (2018). Impossible odds. *Philosophy and Phenomenological Research*. https://doi.org/10.1111/phpr.12517.

Williamson, T. (1990). *Identity and discrimination*. Oxford: Basil Blackwell.

Yablo, S. (1987). Identity, essence, and indiscernibility. *Journal of Philosophy, 84*(6), 293–314.