Reference and Essence

Nathan U. Salmon

PRINCETON UNIVERSITY PRESS
PRINCETON, NEW JERSEY
To my parents,
Mair and Rebecca Ucuzoglu,
without whom *that*(the author of this work) would be possible, but only possible.
in the $K$ and $I$ mechanisms are knowable by a priori "philosophical analysis." Although Kripke might be seen as employing mechanisms similar to the Putnam-inspired $K$ and $I$ mechanisms for generating necessary a posteriori truths, it would be a serious mistake to interpret his use of these modal arguments as an endorsement of the Putnam-inspired program, elaborated by Donnellan and set forth in Chapter Five above. That program uses the $K$ and $I$ mechanisms in an attempt to substantiate the claim that the theory of direct reference has the consequence that there are certain nontrivial examples of a posteriori truths which are nevertheless necessary, and hence, by the theory, essentialist. We have seen that this program is flawed in that it employs as a crucial premise a principle which appears to be quite independent of the theory of direct reference, and which can be made to yield the desired consequence without further help from the theory of direct reference, though neither Donnellan nor Putnam seemed to have recognized this. Kripke, on the other hand, explicitly recognizes the essentialist import of the major premise in his modal arguments generating necessary a posteriori truths. Given this explicit and conscious use of essentialist principles in these modal arguments, one should not interpret Kripke to be offering these modal arguments as substantiation for the claim that nontrivial forms of essentialism are derivable as consequences from the theory of direct reference.

25. Kripke’s "Proof" of the Essentiality of Origin

Though Kripke cannot be interpreted as endorsing Putnam's attempt, as elaborated by Donnellan, to substantiate the claim that nontrivial essentialism is derivable from the theory of direct reference, Kripke does say things that have been interpreted by some philosophers (by myself-from 1973-to 1979, and by others as well) as an endorsement of the claim, with a separate attempt to substantiate it.

25.1. Kripke's Formulation of the Argument

The Putnam-inspired program relied on the use of certain general essentialist principles for its major premise—principles of the form of (27) concerning natural kinds for applications of the $K$-mechanisms, and principles of the form of (37) concerning individuals for applications of the $I$-mechanism. What is left wanting, however, is an argument to show that these principles are themselves consequences of the theory of direct reference. Of course, if a sound argument to this effect can be formed, then this argument alone would substantiate the claim that nontrivial essentialism is derivable from the theory of direct reference, and the $K$ and $I$ mechanisms could be dispensed with. But this seems a dubious prospect.

Kripke also endorses some or all of the crucial essentialist principles employed in the Putnam-inspired program. He says little about how one arrives at the general essentialist principles concerning natural kinds other than that they are knowable by a priori "philosophical analysis." He does, however, say a great deal more about how one may arrive at essentialist principles like the principle (33') involved in the application of the $I$-mechanism to the case of a particular wooden table and its original composition. In fact, in a footnote, Kripke attempts to show that (33'), or something very close to it, can be given something like a demonstration using the direct reference theory of the reference of proper names:

A principle suggested by [my] examples is: If a material object has its origin from a certain hunk of matter, it could not have had its origin in any other matter. Some qualifications might have to be stated (for example, the vagueness of the notion of hunk of matter leads to some problems), but in a large class of cases the principle is perhaps susceptible of something like proof, using the principle of the necessity of identity for particulars. Let 'B' be a name (rigid designator) of a table, let 'A' name the piece of wood from which it actually came. Let 'C' name another piece of wood. Then suppose B were made from A, as in the actual world, but also another table D were simultaneously made from C. (We assume that there is no relation between A and C which makes the possibility of making a table from one dependent on the possibility of making a table from the other.) Now in this situation B ≠ D; hence, even if D were made by itself, and no table were made from A, D would not be B. Strictly speaking, the 'proof' uses the necessity of distinctness, not of identity. The same types of considerations that can be used to establish the latter, can, however, be used to establish the former. . . . In any event, the argument applies only if the making of D from C does not affect the possibility of making B from A, and vice-versa (1972a, p. 114, n. 56).

1This interpretation of Kripke is supported by Albert Cassulo (1977, pp. 152-154).
This passage appears to provide just the sort of argument we have been seeking. For it apparently "proves" a nontrivial essentialist conclusion, using nothing in addition to the theory of direct reference but an innocuous premise asserting that if there is a wooden table $B$ and a separate hunk of wood $C$, then table $B$ could have been made just as it actually is, from the same matter, while a new table $D$ is independently made from the second hunk $C$. What could be less controversial, or freer of nontrivial essentialist import, than this premise?

25.2 Some Initial Considerations

A few points should be clarified before we attempt an analysis of this argument. Kripke's use of the phrase 'something like proof' clearly suggests that he regards the argument as falling somewhat short of a genuine proof of his essentialist thesis. This does not mean, however, that he believes the reasoning to be fallacious or the argument inconclusive. He asserts the argument. He wants to establish the truth of a version of the essentialist thesis (33'), and he clearly intends that the argument be taken as doing just that. One reason he might well balk at calling the argument a proof is that, strictly speaking, it is a derivation from certain assumptions taken as premises, whereas a proof is not. A proof is a derivation from axioms and theorems perhaps, but not from premises. The assumption of the necessity of distinctness might be taken as a theorem of the theory of direct reference, and it might be taken as a premise. (See Chapter One, footnote 35.) For present purposes, let us take it to be a theorem. Even then, we are left with the assumption concerning the possibility of constructing two tables simultaneously from distinct hunks of wood. Though this assumption may be trivial and philosophically uncontroversial, it is not part of a theory of reference. Strictly speaking, it is an independent premise from which the derivation proceeds.

Let us consider what the argument is supposed to show. Kripke intends to derive a special instance of the general essentialist principle mentioned at the beginning of the quoted passage. Specifically, he intends to show that if a wooden table has its origin from a certain hunk of wood, it could not have had its origin in any other hunk of wood. He begins his argument by supposing that we have an arbitrary table $B$ in the actual world constructed from a hunk of wood $A$. Theoretically, however, there is no reason to restrict our initial assumptions to an actual table and actual hunks of wood.

Indeed, it is clear that if Kripke's argument is successful, one can obtain an even stronger conclusion simply by beginning with an arbitrary possible world $w_i$, letting $A$ be the original component material in $w_i$ of some table $B$, whatever kind of material that may happen to be, and letting $C$ be any distinct hunk of matter. Thus we might allow that $A$ be a hunk of wood in $w_i$, whereas $C$ is, say, a sample of water hardened into ice. It must be assumed that $A$ and $C$ are distinct hunks of matter, but they may or may not be hunks of the same kind of matter. If Kripke's specific argument is successful, these more general initial assumptions should yield the stronger conclusion that if a table might have had its origin from a certain hunk of matter, it could not have had its origin in any other hunk of matter. That is, assuming that Kripke's argument is successful, we may similarly derive the strong essentialist thesis that if it is merely possible for a given table to originate from a certain hunk of matter, then it is necessary that the table originate from that hunk of matter and no other. This, of course, is a version of principle (33').

Kripke's argument is perfectly general. Similar considerations can be raised with regard to objects other than tables: other artifacts such as walls and bridges, natural inanimate objects such as mountains and rocks, and even natural organisms such as people. In fact, the argument seems to apply to virtually any sort of object that may be said to have a physical origin and composition. Instead of speaking about the original material from which a given table was made, we may speak of the original gametes from which a given person sprang, and so on. In this way, if Kripke's argument is successful, variants of it may be used to establish several strong essentialist theses concerning the origin and composition of a variety of both animate and inanimate objects. Indeed, as we shall see shortly, a similar argument may even be offered in support of essentialist theses concerning chemical substances, like thesis (22') and analogues of (22') for the special case of chemical elements. For substances may also be said to be composed of more primary or fundamental substances or particulars, namely, component elements in the case of compounds, and atoms having a certain number of protons in the case of elements.

One further point of clarification should be made before we turn to the argument. We shall assume for the sake of simplicity that when Kripke says that a table $x$ was originally made from a hunk of matter $y$, he means that table $x$ was originally constructed entirely from all of hunk $y$, i.e., that no (original) part of table $x$ did not come from hunk $y$, and furthermore that no part of hunk $y$ did not contribute to

198
forming part of table \( x \). It follows from this, presumably, that it is impossible for the same table to be originally constructed from two distinct hunks of matter. In fact, given the appropriate axioms, the supposition that a single table is originally constructed entirely from all of each of two distinct hunks of matter should be a provable contradiction. In any event, we shall take this assumption to be philosophically uncontroversial wherever it is useful to Kripke’s argument. That is, we shall take ourselves to be given the following:

(1) It is impossible for the same table \( x \) to be originally constructed from a hunk of matter \( y \) and in addition to be originally constructed from a \textit{distinct} hunk of matter \( y' \).

25.3. \textit{The Compossibility Premise}

In attempting an analysis of the argument, we have a problem at the outset concerning the special assumption about the possibility of constructing two tables simultaneously. Kripke’s remarks leave it unclear precisely what this premise is. In one place he says that we must assume that the possibility of constructing a table from hunk \( A \) is not dependent on not simultaneously constructing a table from hunk \( C \), and vice-versa. In another place he says that the argument assumes that constructing the very table \( D \) from hunk \( C \) does not affect the possibility of simultaneously constructing the very table \( B \) from hunk \( A \), and vice-versa. (I have altered the wording slightly to make the formulations more perspicuous.) These are quite different assumptions. They are both what might be called \textit{compossibility assumptions}, asserting that two separately possible states of affairs are jointly possible, but the second one is specific in a way that the first is not. The first assumption says only that \textit{some table or other} is to be constructed from each of hunks \( A \) and \( C \), whereas the second assumption tells us furthermore exactly which tables are to be constructed from the two hunks. More precisely, the first assumption Kripke mentions is that if it is possible for some table or other to be constructed from hunk \( A \), and it is also possible for some table or other to be constructed from hunk \( C \), then it is yet a third possibility that some table or other be constructed from hunk \( A \) and in addition some table or other be constructed from hunk \( C \). And the second assumption Kripke mentions is that if it is possible for the very table \( B \) to be constructed from hunk \( A \), and it is also possible for the very table \( D \) to be constructed from hunk \( C \), then it is yet a third possibility that the very table \( B \) be constructed from hunk \( A \) and in addition the very table \( D \) be constructed from hunk \( C \).

Whichever assumption the argument actually employs, it must be seen as placing certain restrictions on admissible choices for hunk \( C \). One trivial restriction, of course, is that it must be possible to construct a table from hunk \( C \). A less trivial restriction is that there not be any matter in common between hunk \( A \) and hunk \( C \). If hunk \( C \) were, for instance, a proper part of hunk \( A \), say its bottom half or an interior portion, the compossibility assumptions may not be satisfied despite the fact that hunk \( A \) and its bottom half are distinct hunks of matter. It may be that this restriction simply reduces to a requirement that \( A \) and \( C \) be nonoverlapping hunks of matter, in the sense that they can have no parts in common throughout their existence. In order to avoid additional complications, let us suppose that it does reduce in this way. Then we may state the two compossibility assumptions Kripke mentions as follows:

(II) For any possible hunks of matter \( y \) and \( y' \), if it is possible for a \textit{table} (meaning \textit{some table or other}) to be originally constructed from hunk \( y \) while hunk \( y' \) does not overlap with hunk \( y \), and it is also possible for a \textit{table} (meaning \textit{some table or other}) to be originally constructed from hunk \( y' \), then it is also possible that a \textit{table} be originally constructed from hunk \( y \) and in addition a \textit{table} be originally constructed from hunk \( y' \).

(III) For any possible tables \( x \) and \( x' \) and any possible hunks of matter \( y \) and \( y' \), if it is possible for table \( x \) to be originally constructed from hunk \( y \) while hunk \( y' \) does not overlap with hunk \( y \), and it is also possible for table \( x' \) to be originally constructed from hunk \( y' \), then it is also possible that table \( x \) be originally constructed from hunk \( y \) and in addition table \( x' \) be originally constructed from hunk \( y' \).

Now, then, which compossibility principle does the argument actually assume? The argument begins by letting \textit{\textit{B}} be a name of an
arbitrary possible table in an arbitrary possible world \( w_i \). We also let ‘\( A \)' name the hunk of matter from which table \( B \) is originally constructed in \( w_i \), and we let ‘\( C \)' name some distinct hunk of matter that also exists in \( w_i \). We want to show that it is impossible for table \( B \) to originate from hunk \( C \), i.e., that there is no possible world in which table \( B \) is originally constructed from hunk \( C \). At this point in his presentation of the argument Kripke invokes the relevant compossibility principle. He writes:

Then suppose \( B \) were made from \( A \), as in the actual world, but also another table \( D \) were simultaneously made from \( C \). . . . Now in this situation, \( B \neq D \).

If we change the description ‘the actual world’ to ‘the world \( w_i \)’ to accommodate our more general initial assumption, it would appear that what Kripke infers from the existence of the first possible world \( w_i \) is that there is an expanded possible world, call it \( w_2 \), in which \( B \) is still a table originally constructed from hunk \( A \), but in which a second table, which Kripke calls ‘\( D' \), is now constructed from hunk \( C \) in such a way that it follows by (I) that the tables \( B \) and \( D \) are distinct. The operative premise in this inference is neither (II) or (III), but something between the two. Unlike (II), it is specific concerning which table is to be constructed from hunk \( A \). But at the point the premise is invoked, there is not yet any particular table which is already constructed from hunk \( C \) and which is to be reconstructed in \( w_2 \). Consequently, unlike (III), the operative premise is nonspecific concerning the second table in \( w_2 \). Thus, given the way in which Kripke’s presentation of the argument proceeds, it would appear that the premise actually used asserts that if it is possible for table \( B \) to be constructed from hunk \( A \), then it is also possible that table \( B \) be constructed from hunk \( A \) and in addition hunk \( C \) be made into a table, some table or other. Given this premise, we simply apply *modus ponens* to infer the existence of a possible world \( w_2 \), in which table \( B \) is constructed from hunk \( A \) and in which a second, and by (I) distinct, table is constructed from hunk \( C \). Once this inference is drawn, we may then name the second table in \( w_2 \) ‘\( D' \).”

*Actually, there is already a problem in supposing that we may give a name to a merely possible table constructed from hunk \( C \), for in order to name something we must first single it out in some way (by description, ostension, etc.). To suppose that we have singled out one particular possible table from all the rest simply by noting that it is (possibly) constructed from hunk \( C \) is to presuppose that there is only one possible table that *could* be constructed from hunk \( C \); i.e., it is to presuppose the principle that we shall call (V) below. (See Chapter One, footnote 41.) This difficulty in naming merely possible objects is a pragmatic difficulty, not a logical difficulty. In fact, instead of taking the letter ‘\( D' \) as a name, we may take it along with the other letters ‘\( A \)', ‘\( B \)', and ‘\( C \)' as *free variables* that occurs within a derivation by way of instantiation. This point does not affect the validity of the argument, since free variables are also rigid (under an assignment of values to variables; see Sections 3.1 and 3.2).
originate from any nonoverlapping hunk of matter. This weaker conclusion does not diminish the significance of the argument for our present inquiry, for even this assertion represents a substantive nontrivial essentiaist thesis. Surely, if such a thesis can be obtained from the theory of direct reference, then the claim that nontrivial forms of essentialism are derivable from the direct reference theory is vindicated.

25.4. The Unfinished Argument

The conclusion that Kripke needs to derive is that table $B$ could not originate from hunk $C$, i.e., that there is no possible world in which table $B$ is originally constructed from hunk $C$. Since $B$ is an arbitrary table in an arbitrary possible world $w_1$ and $C$ is an arbitrary hunk of matter that does not overlap with hunk $A$ ($B$'s original constitutive matter in $w_1$), it will follow from this that if it is possible for a given table to originate from a certain hunk of matter, then it is necessary that the given table originate from that very hunk of matter, or at least from no entirely distinct hunk of matter. The proposition that Kripke actually derives, however, is not that table $B$ could not originate from hunk $C$, but rather the assertion that “even if $D$ were made by itself, and no table were made from $A$, $D$ would not be $B$."

That is, what Kripke explicitly derives is the assertion that there is no possible world in which table $D$ is identical with table $B$, not even a possible world in which table $D$ is made by itself and no table is made from hunk $A$. Although this assertion certainly does follow from the principle of the necessity of distinctness together with (I) and the compossibility premise (IV), it is not yet the desired conclusion. So far we have shown only that in any possible world in which table $D$ is constructed, $D$ still is not the same table as $B$. What we need to show is that in any possible world in which a **table** (meaning any table) is made from hunk $C$, that **table** made from hunk $C$ still is not table $B$.

The situation can easily be represented formally. Let \( T(x,y) \) mean “$x$ is a table that was originally constructed entirely from all of hunk $y$.” We want to show, using only principles from the theory of reference and essentialism-free premises such as (IV), that it is impossible for table $B$ to be constructed from hunk $C$:

\[
(C1) \quad \neg \diamond T(B,C).
\]

What we have so far succeeded in showing is that in any possible world in which table $D$ is constructed from hunk $C$, the tables $D$ and $B$ are still distinct:

\[
(C2) \quad \Box [T(D,C) \rightarrow D \neq B].
\]

The desired conclusion (C1) is, of course, trivially equivalent to the assertion that in any possible world in which some table is constructed from hunk $C$ (and hence in any possible world in which some table is constructed from hunk $A$ and table $B$ is not constructed from hunk $A$), the table constructed from hunk $C$ still is not table $B$:

\[
(C3) \quad \Box (x) [T(x,C) \rightarrow x \neq B].
\]

Conclusion (C3) is formally very similar to our present conclusion (C2). Indeed, it would seem that Kripke's reasoning is aimed in the general direction of something like (C3), since it is equivalent to the desired conclusion (C1). Instead he apparently derives (C2). The issue, however, is not whether table $B$ could be identical with table $D$, but whether table $B$ could be originally constructed from hunk $C$ instead of from hunk $A$. Unless there is some way of moving from our present conclusion to the desired conclusion using only premises that are either entailed by the theory of direct reference or free of nontrivial essentialist import, our ambitious attempt to derive essentialism from the theory of reference is unsuccessful. It would be sufficient to be able to infer (C3) from our present conclusion (C2), since (C3) is equivalent to the desired conclusion (C1). But simply ending the derivation with (C2) is not enough.

The fact of the matter is that (C3), though formally very similar to our present conclusion (C2), simply does not follow from it, not even given all of the direct reference theory. Insofar as the Neat Argument for Kripke's essentialist thesis appeals only to principles from the theory of reference together with premises (I) and (IV), it is simply a non sequitur. If the Neat Argument is going to live up to its acronym, it must rely on some further premise that has not yet been made explicit. This additional premise must for our purposes be free of any nontrivial essentialist import.

25.5. A Principle of Cross-World Identification

How then might we argue from the present premises to the desired conclusion (C1), (C3)? One way would be to make the additional assumption that if it is possible for table $D$ to originate from hunk $C$, then it is necessary that $D$ originate from hunk $C$. It would then seem
to follow that table $B$ does not originate from hunk $C$ in any possible world, since it is table $D$ that originates from hunk $C$ in every possible world, and since $B$ and $D$ are necessarily distinct. But given the arbitrary way in which hunk $C$ and table $D$ were selected, this additional assumption is tantamount to the assumption that if it is possible for a given table to originate from a certain hunk of matter, then it is necessary that the given table originate from that hunk of matter. Since this assertion is precisely what the Neat Argument was designed to “prove,” it obviously cannot be taken as a further premise to the argument. If this were the missing premise, our neat “proof” of essentialism would not only be a failure; it would be a howler.

What Kripke does appear to assume is that in any possible world, any table originating from hunk $C$ is the very table $D$ and no other. More precisely, Kripke appears to assume the following principle as a tacit premise:

(V) If it is possible for a table $x$ to be originally constructed from a hunk of matter $y$, then necessarily, any table originally constructed from hunk $y$ is the very table $x$ and no other.

This principle may be symbolized thus:

$$\Box(x) \Box(y) [(\Diamond T(x,y) \rightarrow \Box(z)(T(z,y) \rightarrow z = x)].$$

One can easily verify that premises (I), (IV), and (V), together with the principle of the necessity of identity and distinctness, yield Kripke’s essentialist conclusion that if it is possible for a given table to originate from a certain hunk of matter, then it is necessary that the given table does not originate from any nonoverlapping hunk of matter:

$$\Box(x) \Box(y) \Box(y') [\Diamond(T(x,y) \wedge y \text{ does not overlap with } y') \rightarrow \Box \sim T(x,y')).$$

We may mimic Kripke’s apparent reasoning as follows. Let $w_1$ be some possible world in which an arbitrary table $B$ originates from some hunk of matter $A$. Let $C$ be any hunk of matter that does not overlap with $A$ in $w_1$. We wish to show that there is no possible world in which table $B$ originates from hunk $C$. Now either it is possible to construct a table from hunk $C$ or it is not. Suppose first that it is not. (Hunk $C$ is absolutely unmalleable, or whatever.) Then it follows directly that there is no possible world in which table $B$ is originally constructed from hunk $C$. Suppose now that it is possible to construct a table from hunk $C$. Then by premise (IV), there is an expanded possible world $w_2$, in which table $B$ originates from hunk $A$, just as in $w_1$, but also in which a second table, which we shall call ‘$D$’, originates from hunk $C$. By premise (I), tables $B$ and $D$ are distinct in $w_1$. It follows by the necessity of identity and distinctness that tables $B$ and $D$ are distinct in every possible world. (Since $B$ and $D$ are distinct in $w_1$, they must also be distinct in the actual world. Otherwise, by the necessity of identity, they would be identical in every possible world, including $w_2$. Since $B$ and $D$ are distinct in the actual world, it follows, by the necessity of distinctness, that $B$ and $D$ are distinct in every possible world.) Now consider an arbitrary possible world $w_3$ in which some table is constructed from hunk $C$. Could that table be the very table $B$ from $w_1$? Given premise (V), it cannot. For by premise (V), the table in question in $w_3$ is none other than table $D$, and $B$ and $D$ are distinct entities in every possible world, including $w_1$. Therefore, there is no possible world in which table $B$ originates from hunk $C$, Q.E.D.

Principle (V) is a crucial component in this piece of reasoning. It tells us that in any possible world, any table originating from hunk $C$ must be $D$ and not $B$. Without some such additional information, there is no reason to suppose that the table in question in $w_1$ could not be $B$. Thus the Neat Argument uses origin as a (necessarily) sufficient condition for being this very table in order to prove that origin is also a (necessarily) necessary condition.\footnote{Recalling a point made in footnote 3, one can see that the theory of direct reference, insofar as it is a theory of closed expressions (proper names, natural kind terms, indexicals, perhaps referentially used definite descriptions, etc.) and not a theory of free individual variables, is entirely inessential to the argument. (Cf. Section 23.) The argument requires only (I), (IV), (V), and the principle of the necessity of identity and distinctness taken in the form

$$\Box(x) \Box(y) [\Diamond(x = y) \rightarrow \Box(x = y)].$$

i.e., as a law of modal logic, not as a special assertion of the direct reference theory of proper names. Our point may thus be put as follows. If the argument from (I), (IV), (footnote continued on next page)
The identification principle (V) is quite compelling. In fact, for many of us, (V) or some weakened version of it may be so fundamental to our point of view that it might easily have escaped our notice at first that the reasoning in the informal presentation of the Neat Argument crucially depends on it. But there is a clear sense in which, given our purposes, any reliance on this additional principle simply begs the question of whether nontrivial essentialism is derivable from the theory of direct reference. For premise (V) is a strong essentialist principle concerning tables and their origins. It asserts that if a given table x is such that it might have originated from a certain hunk of matter y, then the given table x has as an essential property the feature that no table distinct from it originates from hunk y. Looked at in another way, it asserts a nontrivial essential property of any hunk of matter y, namely, that if any table is constructed from it, it is always the same table, or that hunk y has, so to speak, only one potential table “in” it. This additional premise is not something that can be obtained from the theory of direct reference unbolstered by any nontrivial metaphysical theory of essentialism. Since in the end the Neat Argument falls back on a hidden essentialist premise that is quite independent of the theory of direct reference, the argument, taken as an attempt to derive essentialism from the direct reference theory, simply does not succeed.7

and the direct reference theory of proper names to Kripke’s essentialist conclusion is valid without the help of (V), then a similar argument for the same conclusion using only (I), (IV), and the logical law mentioned above is equally valid. (The derivation is the same, except that the letters ‘A’, ‘B’, ‘C’, and ‘D’ are introduced not as proper names but as free variables obtained by universal and existential instantiation.) But it can be proved model theoretically that the latter argument without (V) is invalid. Therefore, the argument using proper names is also invalid without (V).

4More accurately, assuming that intra-world identicals are identical in the absolute, world-independent sense, the principle is simply tantamount to the assertion that any possible table that might have originated from a certain hunk of matter y is in fact the only possible table that could originate from hunk y.

5Cf. Strawson’s remarks concerning the possibility of a different original composition for buildings and ships:

[A particular building]—the Old Bodleian, say—might have been built of [composed of] stone from quarry A instead of being built, as it was, of stone from quarry B.

Will someone say: then it would not have been this building, but another just like it? The retort seems insufficiently motivated. Before the building existed, there existed a plan: a plan for a building on this site, for this purpose, to be constructed of such-and-such type-materials according to such-and-such architectural specifications. . . . If someone said: ‘The QE II, you know, might have been built of quite a different lot of steel from that which it was actually built of’—and gave his reasons—would it not be absurd to reply: ‘In that case it wouldn’t have been the QE II at all—the QE II wouldn’t have existed—it would have been a different ship of that name’? (1979, p. 235).

Strawson is concerned here primarily with an epistemic ‘might have been’, but I have many times heard a similar position taken regarding the metaphysical possibility of a different original composition for certain elaborate or artistic human constructions made according to a pre-determined plan and with certain intentions, etc. The view is often coupled with a rival version of essentialism concerning such objects which locates their essence in their Aristotelian formal, efficient, and/or final causes, rather than in their material cause or in the conjunction of the material cause with one or more of the others. (See, e.g., Wiggins, 1980, p. 125.) Our point is that a proponent of this general position, whether it is coupled with a competing version of essentialism or part of a broader anti-essentialism that denies any nontrivial essence to complex things, will almost certainly want to deny such principles as (V) and its analogues. The theory of direct reference per se is quite independent of these issues.
principle of essentialism nonetheless, and would certainly be rejected by any anti-essentialist. It asserts of any person other than McGinn that he or she could not originate from McGinn’s actual gametes. This thesis, of course, is entirely separate from the theory of direct reference.

Although Kripke never discusses in print how one might establish general essentialist principles of the form of (27) concerning natural kinds, such as (22°) and its analogues, it is easy to see how one might construct an addition argument for the principle that, if a substance S might have had a chemical structure C, then substance S is such that it could not have any chemical structure other than C. We assume the following two premises:

If it is possible for a substance S to have a chemical structure C, and C’ is any chemical structure distinct from C, then it is also possible for substance S to have chemical structure C whereas some other substance S’ has chemical structure C’.

If it is possible for a substance S’ to have a chemical structure C’, then it is necessary that any substance having the chemical structure C’ is substance S’ and no other.

Given these two premises, we reason exactly as before to the desired conclusion. Though it may indeed be by way of this addition argument that one arrives at general essentialist principles concerning natural kinds, the same remarks made above concerning premise (V) of the original Neat Argument apply with equal force to the second premise of this addition argument concerning substances. Like (V), this premise is a nontrivial essentialist principle concerning substances. It is philosophically controversial to the extent that any thoroughgoing anti-essentialist would certainly reject it. Moreover, it is logically quite independent of the theory of direct reference.

As it is stated, (V) (and its analogues) may be too strong. Principle (V) asserts that any table z that originates from the actual constitutive matter of a given table x must be the very table x. It does not require that the table z be constructed in the same way that x is actually constructed, following the same design, or anything of the sort. All it requires is that the table z be constructed from the same matter. It is not at all clear, however, that this is sufficient. Suppose that, in some other possible world w, the constitutive matter of a given table x is shaped into a table that is radically different from x in design and structure. Suppose, for instance, that the portion of matter that actually makes up the top surface of x goes instead to make up the legs of the table in w, and so on. Would the table in w nevertheless be one and the same entity as the original table x, since it is constructed in w from the same matter that x is actually constructed from?9 Principle (V) may be replaced with the following, considerably weaker, assertion:

(V’) If it is possible for a given table x to be originally constructed from a certain hunk of matter y according to a certain plan P, then necessarily any table originally constructed from hunk y according to precisely the same plan P is the very table x and no other.

One may wish to weaken this principle even further by making reference not only to the configuration of the matter of table x but also to such factors as the artisan who constructs x, the artisan’s reasons for constructing x, the time and place of the construction, and so on.) Principle (V’) together with (I) and a correspondingly strengthened version of the original premise (IV) still yield the intended essentialist conclusion concerning tables and their origins.

Principle (V’) is exceedingly plausible, almost to the point of being indubitable. If two tables in two different possible worlds are constructed from the very same stuff in precisely the same way and, let us assume, with exactly the same structure atom for atom, how can they fail to be the very same table? What more could one ask? What more is there to being this very table? The fact remains, however, that even the weaker (V’) is a substantive metaphysical principle that is not entailed by the theory of direct reference. It is supported by a set of intuitions that are entirely separable from our intuitions concerning reference and intensionality.

---

8Similar questions can be raised with regard to McGinn’s principle concerning persons and their origins. Intuitions may differ in the two cases.

9The strengthened version of (IV) that is needed is the following:

For any possible table x, any possible hunks of matter y and y’, and any plan P, if it is possible for table x to be originally constructed from hunk y while hunk y’ does not overlap with y, and it is also possible for a table to be constructed from hunk y’ according to plan P, then it is also possible that table x be originally constructed from hunk y and in addition some table x’ be originally constructed from hunk y’ according to plan P.

If premise (V’) is weakened by including reference to the table’s artisan, time, and place, then premise (IV) must be strengthened accordingly.
25.6. An Alternative Argument

We extracted the Neat Argument from the passage from Kripke quoted in Section 25.1 above. But this is not the only addition argument that can be extracted from Kripke’s presentation of the derivation of his essentialist thesis. Let us consider again the relevant passage, but this time we shall take certain liberties, indicated by brackets, in quoting the argument:

Let ‘B’ be a name (rigid designator) of a table, let ‘A’ name the piece of wood from which it actually came. Let ‘C’ name another piece of wood. [Suppose a table D were made from C and no table were made from A.] Then suppose B were made from A, as in the actual world, but also [the other] table D were simultaneously made from C [as in the second world]. . . . Now in this situation B ≠ D; hence even [where] D [was] made by itself, and no table [was] made from A, D [is not] B.

This is a slightly different argument from the Neat Argument. It starts out the same: We begin by assuming the existence of a possible world \( w_1 \) (or the actual world) in which a table B is constructed from a hunk of matter A, and C is an entirely distinct, nonoverlapping hunk of matter. We wish to show that there is no possible world \( w \) in which table B is constructed from hunk C, i.e., conclusion (C1). At this point the reasoning diverges from the Neat Argument. We derive (C1) not by immediately applying a composibility premise to our assumption of \( w_1 \), but by first making an additional assumption in order to derive (C1)’s equivalent (C2), i.e., that in any possible world \( w \) in which a table is constructed from hunk C, that table is not the very same table as B, but a different table. Assume an arbitrary possible world \( w \) in which an arbitrary table D is constructed from hunk C. We need only show that D ≠ B in \( w \), and we are done. To do this, we invoke a composibility premise like (IV) to construct an expanded possible world \( w_2 \) containing both table B, constructed from hunk A as in \( w_1 \), and table D, constructed from hunk C as in \( w \). By (I), B ≠ D in \( w_2 \). It follows by the necessity of identity and distinctness that B ≠ D in \( w \) — Q.E.D.

This argument is an addition argument, but unlike the Neat Argument, it does not employ the cross-world identification principle (V), or anything as strong, as a premise. It makes do with (I), the necessity of identity and distinctness, and a composibility premise analogous to (IV). In the Neat Argument, (IV) is used to swell the ranks of the tables in the initial world \( w_1 \) by constructing a new, nonspecific table from hunk C. In the present argument, the addition strategy is more specific. In addition to the world \( w_1 \), we assume a second, arbitrary world \( w \) containing a table constructed from hunk C. It is only after this second assumption is made that composibility is invoked, creating a more table-filled world not merely by non-specific expansion of \( w_1 \), but by fusing both of the initial worlds \( w_1 \) and \( w \), with their particular tables B and D, into an expanded world \( w_2 \). This higher degree of specificity in constructing the expanded world \( w_2 \) makes further reliance on the essentialist (V) unnecessary.

Let us call addition arguments employing this more specific addition strategy fusion arguments. Let us call the particular addition argument just presented the Fusion Argument for Tables, or the Fat Argument, for short.

We have already encountered the composibility premise operative in the Fat Argument. It is none other than

\[
(\text{III}) \quad \text{For any possible tables } x \text{ and } x', \text{ and any possible hunks of matter } y \text{ and } y', \text{ if it is possible for table } x \text{ to be originally constructed from hunk } y \text{ while hunk } y' \text{ does not overlap with hunk } y, \text{ and it is also possible for table } x' \text{ to be originally constructed from hunk } y', \text{ then it is also possible that table } x \text{ be originally constructed from hunk } y \text{ and in addition table } x' \text{ be originally constructed from hunk } y'.
\]

If this composibility premise is free of nontrivial essentialist import in the way that (IV) is (or is itself a consequence of the theory of direct reference), then the Fat Argument succeeds, for our purposes, where the Neat Argument failed. For the Fat Argument generates the same essentialist conclusion as the Neat Argument without invoking the latter argument’s extra essentialist premise (V).

Unfortunately (III), unlike (IV), is nontrivially essentialist, in our sense, though perhaps not obviously so. Principle (III) trivially entails, by letting x and x' be the same, that if it is merely possible for a given table \( x \) to originate from a certain hunk of matter \( y \), then it is in fact necessary that the given table \( x \) does not originate from any nonoverlapping hunk of matter \( y' \), provided that it is impossible for the given table \( x \) to be constructed entirely from all of hunk \( y \) and in addition entirely from all of the distinct hunk \( y' \). In fact, this special instance of (III), with \( x \) and \( x' \) the same, is the only instance actually needed in the Fat Argument. Although this premise is slightly weaker than Kripke’s essentialist thesis without the additional provi-
so, it hardly has any less essentialist import. (See Chapter Three, footnote 3.)10 Moreover, it is fairly obvious that this premise is no mere consequence of the theory of direct reference.

There are at least two reasons why premise (III), although nontrivially essentialist, appears at first glance to be no more objectionable than the less specific compoisibility premise (IV). One reason is that the bound variables ‘x’ and ‘x’ occurring in the statement of (III) are distinct, and consequently the reader immediately thinks of instances where these two variables take on distinct entities as values. For distinct tables x and x’, principle (III) is quite plausible, and indeed quite free of any nontrivial essentialist import.

The other reason that (III), though essentialist, might appear metaphysically hollow at first blush is that (III), and (IV) as well, are reminiscent of, and easily confused with, (II), the least specific of the three compoisibility principles. Principle (II) is a perfectly trivial and uncontroversial truth, but by the same token, like (IV), it is too weak to take us from (I) and the necessity of identity and distinctness to nontrivial essentialism.


Loosely speaking, all that separates (III) and (IV) from their nonspecific counterpart (II) is the Neat Argument’s premise (V), the principle that sameness of original constitutive matter is a sufficient condition for the cross-world identification of tables. Principle (II) together with (V) entail both (III) and (IV). In fact, what I want to suggest, roughly, is that nonspecific compoisibility principles like (II) are typically the source of specific compoisibility principles like (III). Whatever plausibility there is to (III) rests largely on the obviousness of (II). If we believe (III), we do so in some sense because we believe (II), and we are prone to make cross-world identifications of the sort licensed by (V) or (V’). Of course, one may believe (III) for other reasons, without basing it on (II) and (V), but the intuitive obviousness of (II) together with a deep-seated inclination to reason in accordance with (V) provide a natural and compelling basis for (III). The idea is something like this: In some sense, we first recognize the obvious truth of (II). Then, making the cross-world identifications, we also assent to the more specific principles (III) and (IV). Thus, (III) and (IV) are, as it were, the epistemic progeny of (II) and (V). Both (III) and (IV) closely resemble their parent (II), but (III) remains nonessentialist like (II), whereas (IV) is essentialist, inheriting its essentialist import from (V).

Of course, the anti-essentialist who rejects (V) and (V’) may still accept (IV) on its own hook, so to speak, or on the basis of some very general anti-essentialist principle, e.g., that anything not provably contradictory is possible (assuming that, given the appropriate axioms, the supposition that the same table is originally constructed entirely from all of each of two distinct hunks of matter is a provable contradiction). But what I am suggesting is that, for many of us, the identification principle embodied in (V’) plays some role in our acceptance of at least (III), if not (IV) as well. Such cross-world identification principles, though typically not explicit, are fundamental to a very common point of view—perhaps the ordinary point of view—concerning what would remain true if a certain portion of matter were made into an object such as a table. These identification principles are so central to the common point of view, so deeply internalized, that it may often completely escape our notice when our modal thinking—as in the informal Neat Argument, and if my speculations are correct, in the Fat Argument as well—depends crucially on just such identifications.11

This is not to say that all such cross-world identification principles are categorically true, or to forget their nontrivial essentialist import. It is only to suggest that they play a basic structural or organizational role in our modal thinking, or at least in the standard modal

10 See also footnote 5 above, which applies mutatis mutandis to the Fat Argument. The argument yields nontrivial essentialism using only (I), the special instance of (III), and the necessity of identity and distinctness taken as a logical law. We are granting that (I) is philosophically uncontroversial. The essentialist import of the logical law is only trivial haecceity essentialism. The nontrivial, substantive essentialism comes entirely from (III).

Kripke has informed me in discussion of these issues that it was in fact the Fat Argument, not the Neat Argument, that he had in mind when he wrote his footnote—though he acknowledges that the footnote is so compressed as to admit an interpretation resulting in the Neat Argument. (He did not use these names, of course. See Salmon, 1979a and the first page of the preface of the book edition of Naming and Necessity, especially its second footnote.) It should be noted again here that Kripke has also informed me that it was probably not part of his intention in his footnote to derive a nontrivial version of essentialism without relying on anything essentialist beyond the theory of direct reference. See Chapter Three, footnote 11.

11 See also Chapter Two, footnote 5.
thinking of many of us. The anti-essentialist philosopher may still consistently deny these principles. In so doing, the anti-essentialist may go against the grain of the way many of us organize our thought about certain counterfactual situations, but so far, it would seem that acceptance of the theory of direct reference per se presents no obstacle whatsoever to the anti-essentialist position. The two positions are perfectly compatible. (Cross-world identification principles like (V') are discussed and scrutinized in Appendix I below.)

Conclusion

The case has not been successfully made that nontrivial forms of essentialism concerning natural kinds or concrete individuals, such as artifacts or organisms, are genuine (modal logical) consequences of the theory of direct reference supplemented only by premises themselves philosophically uncontroversial and free of nontrivial essentialist import. Each of the arguments that appear to derive nontrivial forms of essentialism from the philosophy of language involve in the end some hidden nontrivial essentialist premise or other that cannot be shown—or at the very least, has not been shown—to be itself completely reducible in some way to the philosophy of language. Our essentialist intuitions aside, the theory of direct reference seems perfectly compatible with the claim that examples such as ‘Water is H₂O’, ‘Tigers are mammals’, and ‘This very table originated from hunk of wood H’ presented in Section 8.1 are contingent a posteriori rather than necessary a posteriori. No conclusive reason has yet been given to suppose that the essentialist principles entailing the necessitations of these statements are anything but what they appear to be: theses stemming from an irreducibly metaphysical, and philosophically controversial, theory of essentialism, a theory that is no mere consequence of the philosophy of language.

The theory of direct reference bears directly on several important issues in philosophical semantics. We have seen, for instance, that it entails that proper names are rigid designators, and hence it entails the existence of certain necessary but apparently a posteriori truths, e.g., that Hesperus is identical with Phosphorus. The theory might also entail the existence of certain contingent but a priori truths, e.g., that Neptune is a planet causing perturbations in the orbit of Uranus (a priori for Leverrier). We have seen that the theory even entails a certain trivial form of essentialism, e.g., that Hesperus is such that it could not fail to have the property of being Phosphorus, and hence