Notre Dame Journal of Formal Logic
Volume 31, Number 3, Summer 1990

On Defining Identity

ELIAS E. SAVELLOS*

Abstract I argue that identity must be viewed as an indefinable, primitive notion. I first show that the popular attempts to define identity do not successfully escape the charge of circularity. I then argue that this is no accident. Any attempt to define identity is bound to be circular, since the intelligible understanding of the definition of identity *must* make recourse to the intelligible understanding of identity itself.

1 Since Frege's time the two-place predicate '=' of *identity* has been viewed primarily as a logician's notion. In its Fregean or classical formulation – the formulation that is to be found in our standard logic textbooks – identity theory can be obtained by adjoining to ordinary quantification the axiom schemata:

(A1) $x = x$ (Reflexivity)

and

(A2) $(x = y \;\&\; Fx) \rightarrow Fy.$ (Indiscernibility)

The formula system thus obtained is both demonstrably complete and, as Quine has shown us ([12], p. 180), it has a categorical interpretation: any two introduced predicates that separately satisfy axioms (A1) and (A2) cannot but be coextensive. It is exactly these features of identity theory, Quine tells us ([10], pp. 62–63), that provide us with a powerful reason to reckon identity theory to logic. The universality of identity theory, its indifference as to the values of our variables, provides us with another.

Though well placed within the logician's realm, the status of identity itself has remained a controversial matter. For some, identity is a primitive notion es-

*I would like to thank Mark Bernstein, Jon Kvanvig, and Christopher Menzel for their comments on an earlier draft of this paper.

Received January 3, 1989; revised March 17, 1989

caping all attempts of definition or analysis. Yet attempts to define identity have not been unpopular. Frege himself set the pace for the controversy. In the *Grundlagen*, Frege speaks of the principle '*a is b* just if *a* and *b* have exactly the same properties' as fixing the sense of '='. Yet in his own formal system, Frege presents '=' as primitive. And then he argues that we *must* view '=' as primitive: "Since any definition is an identity, identity itself cannot be defined" ([14], p. 80).

As Dummett forcefully pointed out, since Frege's reasoning is based on a narrow understanding of definition, it is not compelling.[1] On the other hand, if we heed to Frege's *insight* as to the conceptual priority of identity over that of definition, we have the means to resolve the controversy as to the primitiveness of '='. So, at any rate, I will argue in this paper. For I will show that identity must be viewed as a primitive indefinable notion, since any attempt to define identity will turn out to depend upon identity itself. I will first show that the popular attempts to define identity do not successfully escape the charge of circularity. I will then argue that this is no accident. The intelligible understanding of the definition of identity must make recourse to the intelligible understanding of identity itself.

2 Standardly, within a first-order predicate calculus, the two-place predicate '=' of identity is taken as a primitive symbol. This is not to say, however, that there are no exceptions. Notably, Quine has repeatedly insisted that identity need not count as a primitive predicate. At least this is so for any theory *T* that has a finite lexicon of primitive terms. In such a theory, Quine argues ([11], pp. 13–15 and [10], pp. 63–64), we have the means to eliminate '=' in favor of a complex predicate construed by exhaustion of all the other primitive predicates of the theory. In *T* we can *simulate* identity for *a* and *b* as long as *a* and *b* share all the primitive predicates of *T*.

Had Quine's proposal worked, there would be no need to view '=' as primitive. And we would not need to bother with the rather troublesome anomaly associated with the view of '=' as a logical constant: namely that '=' appears to enable us to form atomic sentences (see [3], p. 542 and [10], p. 61). Yet Quine's proposal does not work. Its shortcomings have been adequately pointed out in the literature, and I could do no better here than to summarize what appears to be the main problem[2]: where *T* is a theory about *Ks*, and where the list of primitive predicates of *T* is limited to a finite number *n* we could fail, by Quine's method, to provide for anything beyond indistinguishability in limited respects *n*. On the other hand, adding predicates to the list so that identity for *Ks* is adequately simulated would seem to involve us in the kind of decision that identity conditions for *Ks* provide. Take, for example, material objects as the values of the variables of *T*. Where our stock of primitive predicates of *T* includes *n* predicates except spatiotemporal similarity, we can hardly take indistinguishability in *n* respects for *x* and *y* in *T* to count as identity for *x* and *y* in *T*. By adding spatiotemporal similarity to our list, on the other hand, identity would be sufficiently simulated, yet only because spatiotemporal similarity is exactly what is appropriate for material object identity.

Quine's would-be first-order definition of identity does not work. But when we consider higher-order languages, defining identity seems within reach. In particular, it has been thought that we can fix the sense of '=' by means of the second-order principle (P) below:

(P) $(x)(y)[x = y \leftrightarrow (F)(Fx \leftrightarrow Fy)]$.

Where 'x' and 'y' range over individuals and 'F' ranges over properties, (P) simply says that individuals are identical if and only if they share all their properties.[3] A suitable definition of identity based on (P), Baruch Brody recently argued ([2], p. 68), can be given by (B) below:

(B) $x = y =_{\text{def}} [(F)(Fx \leftrightarrow Fy)]$.

Now as is well known, there is widespread discontent with (P) taken from right to left. Though I could not here go over the various grounds of discontent,[4] it is important for the goal of this paper to remind ourselves of a criticism often made in the literature: to decide that $x = y$, we will have to decide that x and y share all their properties, and this latter decision cannot be made unless we already know in advance that x and y are identical (see, e.g., [3], pp. 544-545 and [14], pp. 49-50 and 55-57). If this criticism of (P) is right, as I believe it is, community of properties as a criterion for deciding the truth of identity statements is beset with circularity.

When we turn to (B), the definition of identity based on (P), we seem to be faced with a similar difficulty: the apparent violation of the noncircularity requirement for an adequate definition. In its most narrow understanding, the requirement is that the definiens be free from occurrences of the definiendum. In a wider sense the noncircularity requirement can be expressed, I believe, thus:

(NCR) The definition of an expression S ought not to depend either directly or indirectly on a prior intelligible understanding of S itself.

The credentials of the noncircularity requirement can be traced as far back as Aristotle's *Topics*, and its almost uniform occurrence both in our logic books as well as in comprehensive theories of definition is embedded in its simple justification (see, e.g., [1], pp. 68-70 and [13], pp. 152-59): the ultimate function of any definition is to introduce an expression S into a language L by specifying the meaning of S in terms of expressions already available in L. To the degree that we draw upon a prior comprehension of the expression S itself, we cannot say that we have succeeded in making S more comprehensible to someone who still does not understand S.

Since the definiens of (B) does not contain the '=', on a narrow understanding of the noncircularity requirement we seem to have no violation. On the other hand, (NCR) seems clearly violated once we agree with Brody that the domain of the quantifier 'F' is the class of all properties. For then identity properties are included in the scope of the quantifier, and identity is defined in terms of itself.

How then does Brody defend (B) in the face of apparent circularity? To be sure, Brody argues ([2], pp. 11-14), the definiens of (B) appeals to a totality of properties and identity is one of them. Yet this, Brody continues, only shows that (B) is an *impredicative* definition and as such it does not violate (NCR) either conceptually or epistemically. We are not threatened by conceptual circularity, for the intelligible understanding of (B) is independent of our intelligible understanding of identity itself. All one needs in order to grasp (B) is to understand what a relation is and what the holding of a relation is, as well as what it is for individuals *a* and *b* to share all their properties. Nor does (B) violate (NCR) epistemically, for there is a way to come to know that *a* and *b* are identical without having to have first checked every property (and thus identity itself):

> We look at some of the properties, see that they have them in common, and infer that they have all of their properties in common—that they are identical. What type of inference is that? Well, if the properties in question include one—other than identity properties—that two different objects cannot share, then the inference in question is a deductive inference. If no such property is included in the set of properties in question, but the set in question is such that it is unlikely that two different objects would share it, then we infer that *a* is identical with *b* as a way of simplifying our explanatory account of things. ([2], pp. 13–14)

Should we grant Brody the claim that (NCR) is not violated by (B)? I do not think so. The argument of the next section will demonstrate that we should not think that (B) could avoid conceptual circularity. Nor should we think that Brody's attempt to defend (B) against the charge of epistemic circularity is successful either. For it is difficult to see how we can employ Brody's suggested method without either ending up throwing the baby out with the bath water, or coming about in a circle: if we take Brody at his word, then on discovering either some property that 'two different objects cannot share' or a set of properties that 'it is unlikely that two different objects would share it', we can immediately and without the aid of (B) infer identity. Surely, whatever the merits of this inferential procedure may be, we hardly have a defense of (B). On the other hand, it is obvious that if we expand the inferential procedure to make a direct appeal to (B), then we had better not rely on (B) taken from left to right. But this is exactly what needs to be done if we are to turn Brody's comments into a relevant suggestion in this context: we go from distinctive property to failure of nonidentity and then to identity, after inferring community of properties from identity.

There is a second point. How exactly do we check on any property *F* of *a*? Doing so implies that we can individuate *a*; that is, we can isolate *a* in experience by drawing its spatiotemporal boundaries and distinguish it from among other *a*s and its environment.[5] Yet this in turn requires the possibility of identity judgments. There are at least two reasons for this: (a) The claim that *a* has been individuated logically requires the ability to answer successfully the question, 'How many *a*s are there?'. But even to begin answering this question we must be able to tell whether we are counting the same *a* twice over. And for this task we need identity. (b) To *know* that *a* has been individuated (in experience) commits us to the possibility of *a*'s reidentification from a variety of perspectives and points of observation. This is not merely to guard against the possibility of illusion. It is a requirement of objective knowledge that further observation *could* be made.[6] But how is this possible without our prior ability to make identity judgments? In the end Brody's attempt does not seem very successful. Just as Russell and Poincare thought, impredicativity hardly escapes the charge of circularity.

Principle (P), the best candidate for defining identity, fails on account of violating (NCR). It is tempting at this point to go on examining other plausible candidates and attempt to show that these too violate (NCR). Yet, on the one hand, the successful accomplishment of such a task, though it would surely reinforce our suspicions, would hardly suffice to demonstrate conclusively that identity is an indefinable, primitive notion. On the other hand, undertaking such a task is not necessary, for, as I would now like to show, a direct proof of the indefinability of identity is available.

3 Let me first present the argument in a synoptic, abstract form. Later I will expand and make it more concrete by means of an example.

Suppose that we want to define some n-place predicate P of a language L. Our definitional attempt would take the form:

(DA) $P(x_1, \ldots, x_n) \leftrightarrow S$

where, standardly enough, the following conditions need to be satisfied: (a) $x_1, \ldots, x_n$ are distinct variables, (b) the definiens has no free variables other than the variables $x_1, \ldots, x_n$ that occur in the definiendum and (c) in the definiens, the only nonlogical constants are either primitive or previously defined symbols of L (see [13], p. 156 and [9], pp. 87–88). Now consider condition (b). As is well known, the formal reason for its postulation is to prevent the occurrence of contradictions that in turn threaten the requirement of noncreactivity of definitions (see [13], Section 8.3, pp. 155–160 and [9], p. 87–88). Yet, it is not enough merely to say that the free variables of the definiens ought to be none other than those of the definiendum. What must be understood is that, on pain of incoherence, the intelligibility of (DA) requires that the free variables of the definiendum *must* also occur in the definiens. What is more, for any free variable x occurring in the definiendum and the definiens, respectively, one ought to understand that in any concrete instantiation of x, *one and the same thing is being picked out.* But this is to say that the intelligible understanding of (DA) rests upon a prior understanding of co-reference. Though in itself innocuous, this result surely hinders the noncircular definition of *identity.*

Let me now bring down to earth the foregoing rather abstract argument: Suppose for simplicity that the n-place predicate we want to define is the predicate 'cat'. Assume further that the meaning of 'cat' is adequately captured by some such expression as 'creature that such-and-such' and let 'G' stand for this latter expression. Our definition of 'cat' more perspicuously articulated is then:

(D) x is a cat $=_{\text{def}}$ x is a G.

Consider now what we do not (and should not) and what we do (and should) understand from (D). (D) does not, of course, say of some specific object that it is a cat just if it is G. Rather, we properly think of (D) as being hypothetical in character. It says that if some object or thing x was a cat, then x would be a G, and vice versa, whatever that thing or object x may be. The variables that occur in each of the expressions that flank the '$=_{\text{def}}$' of (D) are thus not to be thought of as making a specific reference to some particular object or thing. Properly understood, the variables occurring in our definition are nothing but devices playing the role of individual constants. That is, they are devices reinterpretable on occasion as singular terms picking out some specific object or thing.

That the variables occurring in (D) are to be properly thought of as schematic nonreferential devices does not mean, however, that in the context of (D) reference is an alien matter. On the contrary, reference (and indeed co-reference, I submit) is an important conceptual prerequisite of our intelligible understanding of (D). For though the variables of (D) could not in fact *be* co-referential, the intelligible understanding of (D) requires that they are *treated as* co-referential in the following sense: we must think that for 'x' as it appears in the definiendum and 'x' as it appears in the definiens of (D), on any given occasion of their interpretation as singular terms picking out an object or thing, *the very same* object or thing would be the bearer of both of those terms on that occasion. For we certainly cannot understand (D) as saying that if something is a cat, something is a G and vice versa and 'something' picks out a distinct object or thing in each of its occurrences. Rather, intelligibly understood, (D) says that for the object of which we would say that it is a cat, for *that same object* we would say that it is a G. If we say, that is, that a given object *a is* a cat, then for *that object a* we would say that it is a G and similarly for any other object whatsoever. Yet to say that (D) must be sensibly understood in this manner is to say that, on pain of incoherence, our thinking of the pair of variables of (D) as being instantiated is our thinking of them as being instantiated in a co-referential manner. And this is to say that my sensible understanding of (D) requires my understanding of what it is for co-reference to take place.

There is nothing in what has been said thus far that in any way hinders the definition of 'cat'. Nor would any harm result if we repeated the previous argument in regards to the definition of any other arbitrary n-place predicate; any other predicate, that is, except 'identity'. For consider the schema that our attempt to define identity would take. We want to say something like

(I) $\quad x = y =_{\text{def}} C(x,y)$,

where 'C' specifies a condition analytically explicative of '$=$'. Never mind how exactly we cash out the definiens of (I). The point is that our intelligible understanding of (I) presupposes our understanding of the pairwise co-referentiality of the names thought of as exemplifying the variables of (I). That is, where $\langle x,x \rangle$ is the pair consisting of the first members of the pairs flanking the '$=_{\text{def}}$' of (I), on pain of incoherence we must suppose that on any given occasion of the interpretation of its members as names of objects, one and the same object would be picked out by both. And similarly, we must think of the pair $\langle y,y \rangle$ as consisting of the second members of the pairs occurring in the definiendum and the definiens of (I), respectively. But to say that my sensible thinking of my attempted definition of identity requires my understanding of co-reference, is, of course, to say that my definition of identity appeals to or depends upon identity itself. And this means that identity is an indefinable, primitive notion. It seems that Frege's insight was right after all.

4 How can the argument of the previous section be defeated? Two *prima facie* plausible ways come to mind: We can (a) attempt to drive a wedge between identity and coreference, or we can (b) insist that alternative definitional techniques for identity bypass the circularity defect. Let us look at each of these possible objections in turn.

Now for most of us, objection (a) would seem to carry little or no weight. For our understanding of identity just is, we may want to say, our understanding of two referring expressions picking out one and the same thing. On the other hand, how do we convince someone who insists that identity and co-reference are separable and independently intelligible notions? We can show him, I believe, that identity once again has conceptual priority. For to understand co-reference, we may say, is to understand that some *one and the same* object is being picked out by at least *two distinct* acts of reference. But how can we understand all of these things unless we already understand identity?

Now consider objection (b). The argument presented here, the objector might say, is based on a rather narrow understanding of the definitional procedure. Yet even the briefest look in the literature of definition will make it clear that there are many distinct kinds of definition. Why should we suppose that we can extrapolate the results reached here to all of them? Yet unless we do so, we clearly cannot claim that identity is indefinable.

The objection is not compelling. Indeed, there is an almost labyrinthian situation created by the many and various classifications and distinctions of the definitional procedure. But we do not need Ariadne's thread here. For the large majority of available distinctions concerning the kinds of definitions are distinctions based on alternative understandings of the intent or function of a definition. (Consider, for example, such distinctions as *explicit* versus *implicit, real* versus *nominal, stipulative* versus *reportive.*) Yet as such they are not relevant to the argument of Section 3, an argument that depends only on a structural, conceptual feature of the intelligibility of the definitional procedure. To be sure, there are definitional techniques whose general schema deviates from the normative form of (DA), the form upon which my argument is based—*recursive* and *conditional* definitions are examples. But a moment's reflection on the part of the reader ought to make it clear that the point about the conceptual understanding of the variables occurring in such definitions as being co-referential still applies.[7] Identity, I have argued, cannot be defined without violating the non-circularity requirement for an adequate definition. We must conclude that identity should be viewed as an indefinable, primitive notion. But we must be careful here. First, it must be noted that the primitiveness of identity need not thwart the attempt to elucidate discursively the notion in collateral terms.[8] Even when circular, the philosophical attempt to explicate the notion of identity may still be useful in restraining one from entertaining all sorts of erroneous ideas as to its nature.

Secondly, the considerations brought forth here are pertinent to any attempt to define identity, and they are, moreover, independent of the kinds of objects taken as values of the variables in '$x = y$'. We must not, however, suppose that our inability to define identity precludes our ability to define identity for certain kinds of objects. This is not a heterodox point. Suppose that 'cat' is a primitive notion. This does not mean that it cannot be usefully employed to clarify the meaning of, say, 'bob-cat' or 'tiger-cat'. Similarly with identity. We can use our prior, primitive notion of identity to define identity for objects of ontic kind K

(say, persons, events, material objects), for we now say something extra for the objects for which it holds. In principle, the enterprise of saying what identity among *Ks* amounts to or requires is not threatened by the primitiveness of identity.

NOTES

1. Michael Dummett argues: "It is only possible for Frege to say this because he takes the sign of identity to do duty also for the biconditional, which is in turn possible only because he assimilates sentences to names, viz. of truth values; and in any case it seems more natural to take a definition as a stipulation of the interchangeability of two expressions, rather than of the truth of a sentence connecting them" ([3]), p. 543).
2. The inadequacy of Quine's proposal is argued at length by Wiggins in [14] (cf. pp. 199-201) and by Jackson in [8].
3. Some writers label (P) as 'Leibniz's Law' (see, e.g., Dummett [3], and Griffin [6]). In line with present practice, I reserve the term 'Leibniz's Law' for the first-order schema '$(x)(y)\ [x = y \rightarrow (Fx \leftrightarrow Fy)]$'.
4. These are succincty summarized by Wiggins ([14], pp. 55–57). An extensive discussion is similarly found in Griffin ([6], cf. pp. 2–9).
5. I am here following Wiggins' sense of individuation (cf., e.g., [14], p. 5). Of course, there is also an "indirect" sense of individuation, viz. in language or in thought. But this latter sense is not relevant to my argument here.
6. The point about illusion is due to Hampshire ([7], cf. pp. 42–43 and 47–50). For the requirement of alternative observations, see, e.g., Woods ([15], pp. 121–130) and Gottlieb ([5], cf. pp. 90–91).
7. The deviant form of conditional definitions, the annexing of an antecedent hypothesis, detracts nothing from the formal restriction as to the occurrences of the free variables (see, e.g., Suppes [13], p. 165, and Marciszewski [9], pp. 91–92). It remains to be seen how identity would be defined recursively. If nothing else, recursive definitions just are identities. Finally, it goes without saying that the ostensive definition of '=' is not a serious candidate.
8. Indeed Wiggins sees the project undertaken in [14] as constituting exactly such an attempt (cf. [14], e.g., p. 49).

REFERENCES

[1] Ajdukiewicz, K., *Pragmatic Logic*, D. Reidel, Dordrecht, 1975.

[2] Brody, B., *Identity and Essence*, Princeton University Press, Princeton, New Jersey, 1980.

[3] Dummett, M., *Frege: Philosophy of Language*, Duckworth, London, 1973.

[4] Geach, P. and Black, M., eds. and trans., *Translations from the Philosophical Writings of Gottlob Frege*, Blackwell, Oxford, 1952.

[5] Gottlieb, D., "No entity without identity," *The Southwestern Journal of Philosophy*, vol. IX, (1978), pp. 79–95.

[6] Griffin, N., *Relative Identity*, Clarendon Press, Oxford, 1977.

[7] Hampshire, S., "Identification and existence," in Lewis, H. D., ed., *Contemporary British Philosophy*, 3rd ed., Allen and Unwin, London, 1956, pp. 191–208.

[8] Jackson, F., "On property identity," *Philosophia*, vol. 11 (1982), pp. 289–305.

[9] Marciszewski, W., "Definition," in Marciszewski, W., ed., *Dictionary of Logic*, Martinus Nijhoff, The Hague, 1981, pp. 89–96.

[10] Quine, W. V. O., *Philosophy of Logic*, Harvard University Press, Cambridge, Massachusetts, 1977.

[11] Quine, W. V. O., *Set Theory and Its Logic*, Belknap Press, Cambridge, Massachusetts, 1978.

[12] Quine, W. V. O., *The Ways of Paradox*, Harvard University Press, Cambridge, Massachusetts, 1977.

[13] Suppes, P., *Introduction to Logic*, 2nd printing, D. Van Nostrand, Princeton, New Jersey, 1957.

[14] Wiggins, D., *Sameness and Substance*, Harvard University Press, Cambridge, Massachusetts, 1980.

[15] Woods, M., "Identity and individuation," in Butler, R. S., ed., *Analytical Philosophy*, 2nd ed., Barnes and Noble, New York, 1965, pp. 121–130.

Department of Philosophy
510 Blocker
Texas A&M University
College Station, Texas 77843