

*How to Reason About Vague Objects**

PETER VAN INWAGEN
Syracuse University

It is tempting to think that all vagueness is due to language (just as it was once tempting to think that all necessity was due to language). It is tempting to think that if a speaker asserts something, and if what he has asserted is neither definitely true nor definitely false, this must be because the speaker has employed some vague general term like 'rich' or 'heap' or 'near'. Is this right? Does vagueness exist *in sola lingua*, or can vagueness also exist *in re*? Is there vagueness "in the world"? Can there (in some sense) be vague *objects*? I believe that there can be vague objects, and that there are in fact several ways in which this can happen. In this paper I will discuss one of these ways: There could be an occasion (A) on which a speaker refers to an object without making use of any vague general terms; and there could be another such occasion (B); and it could be that it is neither definitely true nor definitely false that the object referred to on occasion A is identical with the object referred to on occasion B. This obviously implies that there can be singular terms X and Y (which neither contain vague general terms nor are defined by means of vague general terms) such that 'X = Y' expresses a proposition that is neither definitely true nor definitely false; these terms, moreover, can be such that the question 'When we are talking about X and Y, how many things are we talking about?' has these features: (a) 'None' is definitely a wrong answer to it, (b) 'Three', 'Four', etc., are all definitely wrong answers to it, and (c) neither 'One' nor 'Two' is either definitely a right or definitely a wrong answer to it.

If all this is true, then it seems fair to say that there are vague objects. And I think all this *is* true. My conviction is based on reflection on cases like the following one.

Suppose we give a man the pure or “Kripkean” proper name ‘Alpha’, and that our act of naming is a clear and unproblematic case of the successful conferring of a name. Suppose that later we give a man the Kripkean name ‘Omega’ (a similar success). Since Kripkean names are not abbreviations for phrases containing general terms, there is no way in which these names can “inherit” vagueness from the language of the namers. Suppose that these two acts of naming are parts of the following longer story. A man receives the name ‘Alpha’; he steps into a science-fictional machine—call it “the Cabinet”—that effects changes in his brain of just such sorts and magnitudes as to create the greatest possible embarrassed hesitation in the making of identity-judgments by someone who holds *your* theory of personal identity: Whatever factor *you* say constitutes the temporal continuity of a person, the Cabinet does its science-fictional best to produce a troublesome borderline case of it. A man, composed, more or less, of the matter that had composed Alpha, steps out of the Cabinet and receives the name ‘Omega’. I contend that you should conclude that the sentence ‘Alpha is identical with Omega’, spoken in the circumstances imagined, would express a proposition that is neither definitely true nor definitely false.¹ Moreover (I contend) the question ‘When we are talking about Alpha and Omega, how many men are we talking about?’ has these features: (a) ‘None’ is definitely a wrong answer to it, (b) ‘Three’, ‘Four’, etc., are all definitely wrong answers to it, and (c) neither ‘One’ nor ‘Two’ is definitely a right or definitely a wrong answer to it.

One feels tempted to say both that this description of the adventures of Alpha and Omega is plausible and that it is incoherent. *I* do, at any rate. I think that if I were left at the mercy of my untutored inclinations, I should let my inclination to say that this description was plausible override my inclination to say that it was incoherent. But I have not been left at the mercy of my untutored inclinations. The late Gareth Evans² has devised an *argument* for the conclusion that descriptions of this sort are incoherent.³

Here is an informal representation of Evans’s argument. Consider Omega, who is supposed to be “indefinitely identical” with Alpha. If Omega is indefinitely identical with Alpha, then Omega has the property: being indefinitely identical with Alpha. But it is false that *Alpha* is indefinitely identical with Alpha, and, therefore, Alpha *lacks* this property. Therefore, by the principle of the non-identity of discernibles, Alpha and Omega are not identical, *contra hyp*.

The central purpose of this paper is to evaluate this argument. I shall begin by restating it. The restated argument will be both closer to Evans’s

statement of it, and, more importantly, will be clearer. It will be clearer because it will not contain the slippery phrase ‘is indefinitely identical with’. Instead of talking of “indefinite identity,” we shall introduce a sentence-operator ‘indef’, which may be prefixed to sentences of any sort—including, of course, identity sentences. ‘Indef’ may be regarded as an abbreviation for ‘it is neither definitely true nor definitely false that’ or ‘there is no fact of the matter as to whether’ or ‘the question whether...has no definite answer’. Now the revised argument. According to the friends of vague identity, it is neither definitely true nor definitely false that Alpha (he who entered the Cabinet) is identical with Omega (he who has emerged from the Cabinet). For short:

$$(1) \quad \text{indef } \alpha = \omega.$$

We refute them by *reductio*. From (1) there follows by the principle of property abstraction

$$(2) \quad \alpha \text{ has } \hat{x} \text{ indef } x = \omega.$$

(The symbol ‘ \hat{x} ’ is pronounced “the property of being an x such that.”) Sentence (2) is simply a somewhat stilted expression of the proposition that Alpha has the property (or quality, attribute, feature, or characteristic) of being a thing of which it is neither definitely true nor definitely false that it is identical with Omega. Some philosophers, on grounds that are not very clear, deny that names as complicated as this could name properties. Real properties, they say, have short names ending in ‘-ness’ or ‘-hood’. These philosophers may read (2) as saying: One of the things that is true of Alpha is that it is neither definitely true nor definitely false of him that he is identical with Omega.

Now it is obviously true that

$$(3) \quad \sim \text{indef } \omega = \omega,$$

which entails, by property abstraction again, and by the principle that any sub-sentence of a sentence may be replaced by a sentence equivalent to that sub-sentence,

$$(4) \quad \sim \omega \text{ has } \hat{x} \text{ indef } x = \omega.$$

And from (2) and (4) there follows by the Principle of the Non-identity of Discernibles

$$(5) \quad \sim \alpha = \omega,$$

“contradicting,” Evans says, “the assumption with which we began, that the identity statement ‘ $\alpha = \omega$ ’ is of indeterminate truth value.” Well, per-

haps there is a contradiction here, but there is no *formal* contradiction: no line in our deduction is the result of prefixing some other line with a negation sign. (Evans gives instructions for turning his argument into one that yields a formal contradiction, but they are far from clear.⁴) Never mind. Evans's argument is trouble enough for the friends of vague identity just as it stands. The friends of vague identity wish to assert (1) without qualification, and they most definitely do not wish to assert (5)—not without the qualification 'indef'. But, if Evans is right, (5) follows logically from (1); that is, (5) *simpliciter* follows from (1). And, it would seem, one is committed to the assertibility of anything one recognizes as a logical consequence of that which one holds to be assertible.

What shall the friends of vague identity say is wrong with Evans's argument? It might occur to them to be suspicious of the two inferences involving property abstraction. Property abstraction is not universally valid, as the "property" version of Russell's Paradox shows. Suspicions are one thing, however, and proof is another. Can we show that the use made of property abstraction in Evans's deduction of ' $\sim\alpha = \omega$ ' from 'indef $\alpha = \omega$ ' is invalid? Or if we cannot *show* that this use is invalid, can we at least show that it would be reasonable for the friends of vague identity—those philosophers who, antecedently to their exposure to Evans's argument, had what they regarded as good reasons for believing in cases of vague identity—to regard it as invalid? Can we at least provide a candidate for the office of "incorrect use of property abstraction by Evans," a candidate that is a reasonable one on the assumption that it is antecedently reasonable to believe in cases of vague identity, and hence reasonable to suppose that there is *some* flaw in Evans's reasoning? (I think that the "Cabinet" case does make it reasonable to suppose—though it does not prove—that there are such cases and hence that there is such a flaw.)

There would seem to be no way to investigate this question other than by giving a precise codification of the rules of inference that should apply in pieces of reasoning involving 'indef', '=', and property abstraction. I shall do this by providing a semantics for the language of first-order logic with identity and property abstraction, supplemented by the constant 'indef'. Or, rather, I shall provide a fragment of such a semantics. (A full semantics even for first-order logic is a lengthy business.) But what I provide will be sufficient to show that Evans's reasoning is invalid according to that fragment. It is also possible to provide a semantics for such a language that confers validity upon Evans's reasoning; the friends of vague identity will, of course, do well not to accept any such semantics.

The semantics I present will be based on two root ideas. First, since certain of the sentences of the language we shall be treating are to be thought of as neither definitely true nor definitely false, we shall need more than two truth-values. (We shall in fact employ three.) Secondly, if identity is indeed vague, then the semantical relation between name and thing named must also be vague. If, for example, 'Alpha' definitely names x , and it is neither definitely true nor definitely false that $x=y$, then it seems inevitable to suppose that it is neither definitely true nor definitely false that 'Alpha' names y . Our semantics must somehow reflect this consequence of vague identity for the naming relation.

A note on what a formal semantics is. A formal semantics is a systematic statement of which sentences and inferences in a given formal language are to count as 'valid' and which as 'invalid'. A formal semantics does not, as some philosophers seem to suppose, explain the meanings of the items peculiar to the vocabulary of the formal language to which it is applied. You already know everything about the meaning of 'indef' that you will learn from me. I am simply going to tell you what in my opinion is valid and invalid when one reasons with 'indef', '=', 'has', and '^', and tell you this systematically. You can agree with me or not. If you do not agree with me, at least you will be able to predict what steps I shall reject in the proofs you construct to show the incoherency of vague identity. In telling you what the valid rules for manipulating 'indef' *et al.* are, I shall not presuppose that there is any vagueness in the world: I will employ classical set theory and I will talk in the normal way about "objects" and informally apply the rules that usually govern counting, identity, and so on. This might be an improper procedure if I thought that there were *no* cases in which the "usual" rules applied (though I think that even then I could defend my procedure *ad hominem*), but of course I don't think that. Consider this analogy. A nominalist claims to be able to show that a platonistic belief in properties is incoherent; perhaps he claims to be able to deduce ' $\sim a$ is a property' from ' a is a property'. A platonist constructs a semantics for property-talk and, in doing so, asserts the existence only of things that satisfy the nominalist's ontological scruples (individuals and their sums, it may be), and shows that, according to this semantics, the nominalist's reasoning is invalid. There would seem to be nothing exceptionable in this procedure, at least provided that the platonist himself accepts the existence of individuals and sums of individuals (and even if he didn't, his strategy might be defensible *ad hominem*). Let us begin.

First, we have a formal language. Its primitives are: *individual constants*, ‘*a*’, ‘*b*’, ‘*c*’, ...; *variables*, ‘*x*’, ‘*y*’, ‘*z*’, ...; *circumflexed variables*, ‘*x̂*’, ‘*ŷ*’, ‘*ẑ*’, ...; two sentential operators, ‘ \sim ’ and ‘indef’; and two two-place predicates, ‘=’ and ‘has’.

It will be noted that our primitive vocabulary is extremely sparse. We have no predicate-letters, not even zero-place ones (that is, we have no sentence-letters); we have no binary sentential connectives; we have no variable-binding operators other than the abstraction operator (in particular, we have no quantifiers). Moreover, we shall count fewer strings of the items of our sparse vocabulary as well-formed sentences than one might, intuitively, expect. The reason for this minimalism is simple: only a very few symbols and constructions figure in Evans’s argument, the evaluation of which is our present concern. Why load ourselves down with semantical apparatus that is irrelevant to the task at hand? It would be a trivial exercise to embed the semantics of the sequel in a semantics for a more powerful language, one that did contain quantifiers and the rest. That is, this would be a trivial exercise provided that a coherent semantics for the more powerful language could be constructed at all. If it could not, that fact would constitute a cogent argument against the coherence of the notion of vague identity. But this cogent argument would not be Evans’s argument. I call the problem of embedding the semantics that follows in a semantics for a more powerful language “trivial” because it could be done *ad hoc*: that is, if we could devise any semantics for the more powerful language, we could always combine that semantics with our semantics for the sparse language simply by stipulating that in the case of any sentence to which the “general” and the “special” semantics assign different values, the special semantics prevails. In theory, combining one semantics with another in this way—the “brute force” method, one might call it—could produce some totally unacceptable results: say, a true disjunction both of whose disjuncts are false. Doubtless, however, such embarrassments could be purged from the combined semantics at the cost of a little tinkering. Whether our “special” semantics could be embedded in a “general” semantics for vagueness—assuming that one exists—in a neat, pleasing, and intuitive way is a question I am not logician enough to investigate. If the only possible embeddings were *ad hoc*, piecemeal, scissors-and-paste jobs, that would certainly count against my treatment of Evans’s argument in the present paper.

From our primitive vocabulary, we construct (one-place) predicates, (property) abstracts, and (closed) sentences.

predicates

1. The result of flanking the identity sign with a constant and a variable is a predicate (an “identity predicate”; the constant is its *term*).
2. The result of prefixing a predicate with ‘ \sim ’ or ‘indef’ is a predicate.
3. Nothing else is a predicate.

abstracts

The result of prefixing ‘ \hat{x} ’ to a predicate containing ‘ x ’, ‘ \hat{y} ’ to a predicate containing ‘ y ’, and so on, is an abstract (the abstract *formed on* that predicate); nothing else is.

sentences

1. The result of flanking the identity sign with (occurrences of) one or two constants is a sentence (an “identity sentence”; the constants are its *terms*).
2. The result of writing a constant followed by ‘has’ followed by an abstract is a sentence (an “ascription sentence”; the constant is its *subject*).
3. The result of prefixing a sentence with ‘ \sim ’ or ‘indef’ is a sentence.
4. Nothing else is a sentence.

With an eye to defining *validity* (of inference) within this language, we specify *models*. A model determines an *extension* and a *frontier* for each abstract, and a *value* for each sentence.

Roughly speaking, a model assigns to each constant an object to be its referent, and specifies what objects that referent is “indefinitely identical” with. Because our language is so very simple, such an assignment suffices to determine an extension and frontier for each abstract (and a value for each sentence). We are to think intuitively of each abstract as denoting (relatively to a model) a property; the extension of an abstract is, intuitively, the class of objects that definitely have that property, and the frontier of an abstract is the class of objects that neither definitely have nor definitely lack that property. We proceed as follows.

A *universe* is a non-empty set of objects. A *pairing on* a universe is a (possibly empty) set of two-membered sets (pairs) of members of that universe. (These are to be “genuinely” two-membered: $\{x, x\} [= \{x\}]$)

cannot be a member of a pairing.) If x and y , $x \neq y$, are members of a pair (belonging to a certain pairing) they are said to be *paired* (in that pairing).

A *model* consists of a universe, a pairing on that universe, and an assignment of one object in that universe to each individual constant (the constant's *referent* in that model). It will occasionally be useful to call the objects with which the referent of a constant is paired the *fringe-referents* of that constant. (The objects with which an object is paired are to be thought of as the objects such that it is indefinite whether that object is identical with them, and the fringe-referents of a constant are to be thought of as the objects such that it is indefinite whether that constant denotes them.)

The extensions and frontiers of abstracts (in a given model) are determined as follows (as a preliminary step, extensions and frontiers are assigned to predicates):

1. The extension of an identity-predicate contains just the referent of its term; the frontier of an identity predicate contains just the fringe-referents of its term.
2. The result of prefixing ' \sim ' to a predicate having extension e and frontier f is a predicate having extension $U-(e \cup f)$ —where U is the universe of the model—and frontier f .
3. The result of prefixing 'indef' to a predicate having frontier f is a predicate having extension f and an empty frontier.

The extension and frontier of an *abstract* are the extension and frontier of the predicate on which it is formed.

A model also assigns a value to each sentence. A value is one of the three numbers 1, $\frac{1}{2}$, and 0—intuitively, (definite) truth, neither definite truth nor definite falsity, and (definite) falsity:

1. The value of an identity sentence is:
 - 1 iff something is the referent of both its terms
 - $\frac{1}{2}$ iff nothing is the referent of both its terms and the referents of its terms are paired
 - 0 otherwise; that is, if and only if nothing is the referent of both its terms and their referents are not paired

Intuitively, an identity sentence has the value 1 if the referents of its terms are identical, $\frac{1}{2}$ if the referents of its terms are indefinitely identical, and 0 otherwise.

2. The value of an ascription sentence is:
- 1 iff The referent of its subject belongs to the extension of its abstract
 - $\frac{1}{2}$ iff The referent of its subject does not belong to the extension of its abstract, and either (a) the referent of its subject belongs to the frontier of its abstract, or (b) a fringe-referent of its subject belongs either to the extension or the frontier of its abstract.
 - 0 otherwise; that is, if and only if neither the referent nor a fringe-referent of its subject belongs either to the extension or the frontier of its abstract.
3. The values of sentences starting with ‘ \sim ’ and ‘indef’ are determined by the table:

ϕ	$\sim\phi$	indef ϕ
1	0	0
$\frac{1}{2}$	$\frac{1}{2}$	1
0	1	0

An *inference-form* is a finite, non-empty sequence of sentences; the final sentence of an inference-form is its *conclusion* and the earlier sentences its *premises*.

A *counterexample* to an inference-form is a model that assigns a value to its conclusion that is less than the least value that that model assigns to any of its premises. A *valid* inference-form is one that has no counterexample: A valid inference-form is 1-preserving and cannot lead from $\frac{1}{2}$ to 0.

Let us now return to Evans’s argument. If we consider only its premise and its conclusion, we can regard it as an instance of the inference-form

$$I_0 \quad \text{indef } a=b \mid - \sim a=b.$$

We should, of course, expect this inference-form to be invalid; any model whose structure is intuitively that of the “Cabinet” example should be a counterexample to it. Consider the (partial) model

$$\{A, B\}, \{\{A, B\}\}, 'a' \text{ ref } A, 'b' \text{ ref } B,⁵$$

which is so structured. This model—let us call it M —is a counterexample to I_0 : it assigns the value $\frac{1}{2}$ to ‘ $a=b$ ’ (since the referents of ‘ a ’ and ‘ b ’ on M are paired), and hence assigns a value of 1 to ‘indef $a=b$ ’ and a value of $\frac{1}{2}$ to ‘ $\sim a=b$ ’.

Since M is a counterexample to I_0 , it must be a counterexample to at least one of the four inference-forms that Evans's reasoning comprises:

- $$\begin{array}{ll}
 I_1 & | - \sim \text{indef } b = b \\
 I_2 & \text{indef } a = b | - a \text{ has } \hat{x} \text{ indef } x = b \\
 I_3 & \sim \text{indef } b = b | - \sim b \text{ has } \hat{x} \text{ indef } x = b \\
 I_4 & a \text{ has } \hat{x} \text{ indef } x = b, \sim b \text{ has } \hat{x} \text{ indef } x = b | - \sim a = b.
 \end{array}$$

The model M is not a counterexample to I_1 , which is obviously valid: ' $b=b$ ' will have the value 1 on any model, and hence ' $\text{indef } b=b$ ' the value 0, and ' $\sim \text{indef } b=b$ ' the value 1.

The model M is not a counterexample to I_2 . M assigns to ' $x=b$ ' the extension $\{B\}$ and the frontier $\{A\}$; hence it assigns to ' $\hat{x} \text{ indef } x=b$ ' the extension $\{A\}$; the referent of ' a ' on M is A , and hence the value of ' a has $\hat{x} \text{ indef } x=b$ ' on M is 1. In fact, it is easy to see that I_2 is valid. If a model assigns 1 to ' $\text{indef } a=b$ ', it must assign distinct referents to ' a ' and ' b ' and must pair them. But the extension of ' $\hat{x} \text{ indef } x=b$ ' in any model comprises just the things that model pairs with the referent of ' b '.

The model M is, however, a counterexample to I_3 . As we saw in our examination of I_2 , the extension of ' $\hat{x} \text{ indef } x=b$ ' on M is $\{A\}$; the referent of ' b ' is B , which does not belong to this extension; but A is a fringe-referent of ' b ' on M , since M pairs A and the referent of ' b '; hence, ' b has $\hat{x} \text{ indef } x=b$ ' has the value $\frac{1}{2}$ on M ; as does its negation; but, as we saw in our examination of I_1 , the value of the premise of I_3 on M is 1; hence I_3 is invalid.

It follows from the values established in the examinations of I_0 , I_2 , and I_3 that M is not a counterexample to I_4 . But I_4 does have counterexamples. I shall consider them in due course. They are, in a sense I shall briefly outline, of far less importance than our counterexample to I_3 . This is because all of the counterexamples to I_4 assign a value of $\frac{1}{2}$ to its premises and a value of 0 to its conclusion. Thus, while I_4 can lead from $\frac{1}{2}$ to 0, it is 1-preserving. Since the friends of indefinite identity assign a value of 1 to ' $\text{indef } \alpha = \omega$ ', Evans's argument will constitute a *reductio* of their position unless it employs some inference-form that is not 1-preserving. But it is just this feature that our semantical fragment ascribes to I_3 and denies to I_2 , and I_3 , and (as we shall see) to I_4 . The only real interest in the fact that I_4 (unlike I_1 or I_2) can lead from $\frac{1}{2}$ to 0, therefore, is contained in this question: Does that fact tend to render our semantical fragment implausible? I shall take up this question later, and I postpone consideration of the counterexamples to I_4 till then.

Evans's argument, therefore, employs an inference-form that is not only invalid on the semantical fragment I have proposed but which fails to preserve "definite truth." I would make these observations.

(i) One could, of course, construct a semantics that confers formal validity on Evans's reasoning. (It would suffice to replace our rule for determining the value of an ascription sentence with the following rule: ' a has $\hat{x}Fx$ ' has the same value as ' Fa '.) The friends of vague identity will, of course, be understandably hostile to the proposal that they adopt such a semantics. Insofar as there is such a thing as "the burden of proof," it would seem to weigh on the enemies of vague identity: *they* are the ones trying to prove something, *viz.* that the notion of vague identity is incoherent. The friends of vague identity are not trying to prove that vague identity *is* coherent, but only to undermine the appeal of Evans's argument; to do that (I should think) they need only construct a reasonably plausible semantics according to which Evans's reasoning is formally invalid. They need not show that the semantics they have constructed is "superior" (in any sense) to a semantics that endorses his argument, much less "correct."

(ii) The individual components of the semantics I have presented seem to me to be very plausible. And each of them was chosen for its intrinsic plausibility. These components were not chosen, either individually or collectively, with any end in view. The invalidity of the inference-form I_3 and the consequent formal invalidity of the move from step (3) to step (4) in Evans's argument was in no way "engineered." This is simply what popped out of the semantical engine when Evans's argument was dropped into the hopper. I was rather surprised. I should not have been surprised if *both* of the inferences involving property abstraction had turned out to be formally invalid. That one of the inferences should be valid and the other invalid, however, was a quite unanticipated result.

(iii) The only component of our semantics that could reasonably be called into question is, I think, the rule for determining the values of ascription sentences. Let me, therefore, say something to provide some intuitive motivation for this rule.

Suppose that it really is indefinite whether x is identical with y . Suppose that y definitely has the property F . Can it really be definitely false that x has the property F ? In order to focus our intuitions about this matter, let us consider two cases of indefinite identity, a "diachronic" case and a "synchronic" case.

We already have a diachronic case of indefinite identity at our disposal: the case of Alpha, Omega, and the Cabinet. (We may say that we have a diachronic case of indefinite identity when we have an x and a y such that it is indefinite whether $x=y$ and there is a time such that it is definitely true that x exists at that time, and another time such that it is definitely true that y exists at *that* time, and there is no *one* time such that it is definitely true that x exists at that time and definitely true that y exists at that time.) Now suppose that soon after Omega emerges from the Cabinet, we hang him; that is, we hang *him*: it is quite definitely true of Omega that *he* dies by hanging. Could it be definitely false of Alpha that *he* dies by hanging? It is hard to see how this could be, given that it is not definitely false that Alpha is numerically distinct from Omega.

We shall say that we have a synchronic case of indefinite identity if it is indefinite whether $x=y$ and there is a time such that x and y both definitely exist at that time. A “commissurotomy” case, if it were properly constructed, might provide such an example of the indefinite identity of persons. Or we might imagine a two-headed giant, a giant with two brains. Imagine that increasingly elaborate neural connections are established between the two brains till a point is reached at which the two-headed giant has, definitely, *one* brain. It seems plausible to suppose that at *some* point in the establishment of this sequence of connections, there is a person x and a person y such that it is indefinite whether x is y . Suppose that this is the case and that y definitely has a certain property—wisdom, say. Could it be that x definitely lacks wisdom? It is hard to see how this could be, given that it is not definitely false that x is numerically distinct from y .

These examples suggest that if a constant k definitely denotes something x , and there is a y such that it is indefinite whether $x=y$, and y definitely has the property denoted by the abstract F , then ‘ k has F ’ should receive a value of at least $\frac{1}{2}$. (It should obviously receive the value 1 if x definitely has—“in its own right,” so to speak—the property denoted by F .) Or, at any rate, this seems plausible in cases in which F denotes a reasonably “ordinary” property like wisdom or ending on the gallows. Should matters be different if F contains the symbols ‘=’ and ‘indef’? I do not see why they should. These reflections seem to me strongly to support the rule our semantical fragment follows for assigning values to ascription sentences.

Interestingly enough, there is a sense in which one could regard the argument of the preceding several paragraphs as a rather weak one and still find it convincing. More exactly, suppose that one were not thinking in terms of three values, but only in terms of truth and falsity. And suppose

that one were wondering whether to call ‘Tom has wisdom’ true or false, given that Tom is foolish but is indefinitely identical with Tim, who is wise. Suppose that one regarded the sort of considerations raised in the preceding paragraphs as providing a *very weak* reason for saying that ‘Tom has wisdom’ was true; a reason that was, perhaps, plainly overridden by the fact that Tom is foolish, but still a reason that was there to be overridden. Having made these suppositions, consider the following general strategy for distributing the *three* values “definitely true,” “definitely false,” and “neither” among the members of some class of sentences:

Assign “definitely true” to a sentence only if there is *no* reason, however weak, that militates against calling that sentence “true”; and similarly for “definitely false” and “false”; in all other cases, assign the value “neither.” In other words: assign the value “neither” to a sentence if there is the least excuse for doing so.

(This “strategy” is to be looked upon as a rule of thumb or heuristic to be kept in mind by one who is formulating a three-valued semantics—a semantics that assigns *only one* value besides definite truth and definite falsity—, whatever sort of sentences that semantics applies to.) Now one who accepts the thesis that the considerations adduced in the preceding paragraphs provide a reason (however weak) for assigning truth to ‘Tom has wisdom’ if the foolish Tom is indefinitely identical with the wise Tim, and who accepts the proposed strategy, will obviously be moved to accept our “ascription” rule. It seems to me that both the thesis and the strategy are quite plausible.⁶

If we find these reflections congenial, we shall not find it counterintuitive to suppose that Omega does not *definitely* lack the property of being a thing such that it is indefinite whether that thing is Omega. Omega, of course, does not definitely *have* that property; but Omega is *sort of* identical with someone who *does* definitely have that property; and, therefore, Omega *sort of* has the property of being a thing such that it is indefinite whether that thing is Omega. And, therefore, it does not follow from the truth of ‘It is not indefinite whether Omega is Omega’—*or* from the truth of ‘It is not indefinite whether Omega has the property of being identical with Omega’—that Omega lacks the property of being indefinitely identical with Omega.⁷

If Omega *sort of* lacks the property of being indefinitely identical with Omega, it follows that Omega *sort of has* the property of being indefinitely identical with Omega. That is, Omega *sort of* has the property of being a

thing such that it is neither definitely true nor definitely false that that thing is Omega. Does it follow that Omega sort of has the property of being such that it is neither definitely true nor definitely false that it is identical with *itself*? One would hope not, for it is plausible to suppose that this property is flatly self-contradictory and, therefore, is quite definitely not possessed by any object whatever. Now the question whether this does follow can't really be raised in connection with our semantical fragment, because we have no way of representing the property we are worrying about in our simple language. But that is easily enough remedied. Let us simply add to our recursive definition of "predicates" in our formal language a codicil that says that the result of flanking the identity-sign with two occurrences of the same variable is a predicate (a "self-identity predicate"). Then our language will contain the abstract ' \hat{x} indef $x=x$ ', which, intuitively, should denote the property we are interested in. What sort of addition to our semantics should we make to accommodate the items we have added to our language by adding self-identity predicates to our basic stock of one-place predicates? What we need to do is to stipulate an extension and frontier (relative to each model) for every self-identity predicate. And only one stipulation seems even remotely plausible: the extension of a self-identity predicate in a model is the universe of that model; the frontier of a self-identity predicate is empty in every model. This stipulation has the obvious consequence that ' \hat{x} indef $x=x$ ' and its alphabetic variants have an empty extension and frontier in every model, which entails that every ascription sentence containing any of these abstracts has the value 0. Therefore, the inference-form

$$\text{indef } a \text{ has } \hat{x} \text{ indef } x=a \mid - \text{ indef } a \text{ has } \hat{x} \text{ indef } x=x$$

is invalid. Any model in which the referent of ' a ' is paired with something will be a counterexample: the value of the premise will be 1 in such a model, and the value of the conclusion is 0 in every model.

I leave open the question whether a similar result could be achieved in a "natural" way by a properly general semantics for a language containing '=', 'indef' and property abstraction. By a properly general semantics for such a language, I mean one that begins in the usual way by constructing models that provide extensions for two-place predicates formed by flanking the identity sign with variables, and which, as a consequence, generates "automatically," as special cases, extensions for the two kinds of one-place predicates containing the identity sign that have in this paper been treated separately, *viz.*, identity predicates and self-identity predicates.

It is reasonably evident from the foregoing that the friends of vague identity need not be troubled by Evans's argument, considered simply as a piece of formal reasoning. They need only adopt a system of rules for reasoning about vague identities that is consistent with the semantical fragment I have presented and declare that (in their opinion) Evans's reasoning is invalid. If they are wise, they will not claim to have *shown* that Evans's reasoning is invalid.⁸ But there is no reason for them to make so strong a claim. Suppose A says to B, "Your position is incoherent," and backs up this charge by the formal deduction of some sort of incoherency from B's position, a deduction that employs plausible but not wholly uncontroversial principles of reasoning. B need not defend himself by *showing* that A's reasoning is invalid. He need do nothing more than present a system of reasoning that *he* accepts, according to which A's reasoning is invalid. Since A is the one who announced that he was going to prove something, it is up to him to show that his system of reasoning is right and B's alternative system wrong. Or, at any rate, the next move is A's. If A doesn't believe that he can *prove* that his own system is correct, he may try to cast philosophical doubts of one sort or another on B's proposed alternative; if A can do this—and if no corresponding doubts can be raised about his own system—this will constitute a dialectically acceptable reply on A's part. What would certainly be dialectically improper would be for A to insist that B present some sort of *proof* that B's system is the correct one.

In the remainder of this paper, I shall consider four philosophical objections that might be brought against the semantical fragment that I have proposed—and, derivatively, against the judgments it makes about the validity of various inference-forms.

Objection One

You employ a three-valued logic. But how are we to interpret the value '½'? The answer is: We can't. The idea of a *tertium datur* makes no sense.

Reply

For any sentence ϕ , the truth-value of *indef* ϕ must be a function of some feature of ϕ (or of the proposition that ϕ expresses; important as this distinction is, I will ignore it): if *indef* ϕ and *indef* ψ are both true, this must be because of some feature common to ϕ and ψ , just as is the

case with $\sim\phi$ and $\sim\psi$ (common feature falsity) and $\Box\phi$ and $\Box\psi$ (common feature necessity). The feature of ϕ in virtue of which $\text{indef } \phi$ is true is incompatible with the truth *and* with the falsity of ϕ , since $\text{indef } \phi$ is false if ϕ is either true or false. Therefore, this feature may reasonably be called a ‘truth-value’, since it “competes with” truth and falsity. And, therefore, if $\text{indef } \phi$ makes sense at all, there must be a third truth-value, which I designate ‘ $\frac{1}{2}$ ’.

But if it seems just too paradoxical to posit a *tertium datur*, we are not forced to do so. We can read ‘1’ as expressing a feature of sentences called “definite truth” and ‘0’ as expressing “definite falsity” and ‘ $\frac{1}{2}$ ’ as expressing a *tertium* in addition to definite truth and definite falsity. And we could say that these three values are the appropriate ones to employ in investigating validity in a language that contains ‘indef’. In the matter of the relation between these three values and the two classical truth-values, there are two positions whose consequences might be investigated: it might be said that definite truth implies truth and definite falsity implies falsity, but not *vice versa*; or it might be said that, strictly speaking, classical truth and falsity do not exist but are idealizations that it is convenient to employ in cases in which indefiniteness is absent (as in mathematics) or can safely be ignored.

An investigation of these alternatives would take us far afield. The important thesis for present purposes is this: ‘ $\frac{1}{2}$ ’ makes just as much sense as ‘indef’, owing to the fact that it represents that feature of a sentence ϕ in virtue of which $\text{indef } \phi$ is true (or definitely true). The possibility remains, of course, that ‘indef’ does *not* make sense. But, if I understand Evans, he is willing to grant for the sake of argument that ‘indef’ makes sense in order to show that the sense it makes is incompatible with its application to identity-sentences. We might compare his strategy with the strategy of Quine, who (in one place) is willing to grant for the sake of argument that ‘necessarily’ makes sense in order to show that the sense it makes is incompatible with its application to open sentences. (We might, in fact, distinguish two “grades of indefinitional involvement,” the first represented by ‘indef John is rich’ and the second by ‘indef John = James’. The first can be accounted for by the Linguistic Theory of Vagueness—the theory that indefiniteness of truth-value is entirely a product of vague general terms—and the second cannot, just as the first and second grades of modal involvement can be accounted for by the Linguistic Theory of Necessity and the third cannot.) I have not undertaken the task of showing that ‘indef’ makes sense, but only the task of showing that Evans has not

succeeded in forcing the friends of indefinite identity to admit that their position faces a special sort of incoherency that they do not face as friends of indefiniteness *simpliciter*.

Objection Two

“Either ‘Omega’ denotes the man who entered the Cabinet or else it does not. That is, in performing your act of naming, you either succeeded in giving the name ‘Omega’ to the man who entered the Cabinet or you didn’t. Moreover, the man who entered and the man who emerged both exist. They’re *there* (timelessly speaking). And either ‘they’ are identical or they are not.”

Reply

This would seem to be either a simple denial of the possibility of vague identity, or else an argument whose premise is the Law of the Excluded Middle. In the former case (I suppose), one could counter it with a simple *assertion* of the possibility of vague identity. As to the Law of the Excluded Middle, this “law” has no legal force at all in conversations in which ‘indef’ figures. It is natural to suppose that a disjunction has the value that is the greatest value of any of its disjuncts. On that natural assumption, if ‘John is rich’ has the value $\frac{1}{2}$, so will

John is rich $\mathbf{v} \sim$ John is rich.

And the same goes for

Alpha = Omega $\mathbf{v} \sim$ Alpha = Omega.

Since, one would suppose, to accept a proposition is to accept it as true (or as definitely true, if a distinction is made between truth and definite truth), one who holds that ‘Alpha = Omega’ has the value $\frac{1}{2}$ is thereby debarred from accepting the proposition that Alpha = Omega $\mathbf{v} \sim$ Alpha = Omega. (Exercise for the reader: show that $\sim \text{indef } \phi \mid - \phi \mathbf{v} \sim \phi$.) I am aware that it is possible to construct a three-valued logic according to which the alternation of a sentence with its negation always has the value 1. Here is how to do it. Rename what we have been calling the ‘value’ of a sentence the “supervalue” of a sentence.⁹ Say that the supervalue of a disjunction is equal to the greatest supervalue had by any of its disjuncts. And say that the *value* of a sentence is the supervalue had by that sentence “under all resolutions of vagueness” if there is such a

unique supervalue, and is otherwise its actual supervalue. For example, the value of ‘John is rich $\vee \sim$ John is rich’ is 1 according to this proposal, because no matter what *precise* term replaced ‘rich’ in this sentence, the resulting sentence would have the supervalue 1. This clever device seems to me to answer to no philosophical need.¹⁰ The word ‘rich’ is vague, and the sentence ‘John is rich $\vee \sim$ John is rich’ does not mean ‘If one were arbitrarily to fix a precise boundary between being rich and not being rich, John would fall on one side or the other of it’. If it is neither definitely true nor definitely false that John is rich, then it is not definitely true that John is either rich or not rich. If this *were* definitely true, then it would be definitely true, true without qualification, that either Tom (who contends that John is rich) or Tim (who contends that John is not rich) was right. But it is not definitely true that either Tom or Tim is right. It is, of course, definitely true that *if* there were a sharp boundary between the possession and the non-possession of riches, then it *would be* definitely true that one or the other of them was right. But there isn’t and it’s not. Or so *I* think. But suppose I’m wrong. This will be cold comfort for the enemies of vague identity. It will not suffice for their purposes that the sentence ‘ $\alpha = \omega \vee \sim \alpha = \omega$ ’ express a definite truth; not unless one can infer from the definite truth of this disjunction that if Tom says, “ $\alpha = \omega$ ” and Tim says, “ $\sim \alpha = \omega$,” then, definitely, one of them is right. If the Law of the Excluded Middle does not warrant this inference, that law is of no use whatever to the enemies of vague identity.

Objection Three

“But how many people *are* there in the Alpha-Omega story? How many members has the set {Alpha, Omega}? You have no coherent way of answering this question.”¹¹

Reply

That this set is empty is definitely false. That it has one member is not definitely true or definitely false. That it has two members is not definitely true or definitely false. That it has one member *or* has two members is not definitely true or definitely false. That it has one member *and* has two members is not definitely true or definitely false. That it has three or more members is definitely false.

[‘That’s crazy. A set has to have a definite number of members, a definite cardinality. That’s essential to the idea of a set.’]

I’m not sure of that. I concede that it’s a theorem of set theory. But perhaps set theory is an idealization. I am enough of a platonist about mathematical objects to take seriously the idea that while there really are sets, they don’t have precisely the properties that set theory ascribes to them—just as, while there really are gases, they don’t have precisely the properties that the kinetic theory of gases (an idealization) ascribes to them. But if you insist that every “set” has a definite number of members, then I shall say that there is no such set as {Alpha, Omega}.

[‘There has to be such a set, since there are such *individuals* as Alpha and Omega.’]

Alpha and Omega are individuals in the metaphysical sense, I suppose. But ‘individual’ is a technical term in set theory: individuals are non-sets that are capable of being members of sets. If “being a set” entails “being an object that perfectly conforms to the requirements imposed upon ‘sets’ by what is called ‘set theory,’” then it is clear that if x and y are such that it is neither definitely true nor definitely false that they are identical, then x and y are not “individuals” in the required sense, owing to the fact that every set that perfectly conforms to the requirements of set theory has a definite number of members. If set-theorists sometimes define individuals as non-sets (or non-classes, if “class” is distinguished from “set”), that is only because they either reject or have not considered the thesis that there is such a thing as vague identity.

[‘You yourself, in constructing your semantics, have employed set-theoretical notions. Universes, pairings, extensions, and frontiers are all sets.’]

I have employed these notions only in giving a formal semantics. Recall my analogy of a formal semantics for a platonistic language whose universe of discourse consisted entirely of objects acceptable to the nominalist. There is no reason that a pure, formal semantics for a language containing ‘indef’, ‘=’ and ‘∩’ cannot be constructed whose universe of discourse comprises only objects among which no indefinite identities hold (numbers, say, or men who have not got involved with the Cabinet). Suppose that the fragment of a semantics I have offered does have a universe of discourse comprising only “definite” objects. It nevertheless succeeds in distinguishing “valid” from “invalid” inferences, and its use of classical set-theoretic notions is unobjectionable. If we move from a “pure” to what Plantinga has called a “depraved” semantics, however, we shall have to

be more careful about our use of set-theoretic notions. If we really do propose, in our semantics, “in the meta-language” as they say, to quantify over objects among which indefinite identities hold, then we shall have to take one of two courses. If we want it to be literally true that there are such *objects* as universes and pairings and frontiers and so on, then we shall have to resort to a set theory that allows sets of indefinite cardinality. (An example of a set of indefinite cardinality would be a set of which it is definitely false that it is empty, neither definitely true nor definitely false that it has one member, neither definitely true nor definitely false that it has two members, and definitely false that it has more than two members.) As far as I know, no such theory has actually been constructed. (So-called “fuzzy” sets are not sets of indefinite cardinality. In fuzzy set theory, one has various “degrees” of set-membership—perhaps as many degrees as there are real numbers between 0 and 1—, and to specify a set is to specify, for each object, what its degree of membership in that set is. Classical set theory is a special case of fuzzy set theory, the case in which there are just two degrees of membership: “Yes” and “No,” one might call them. A fuzzy set does not really have such a thing as a cardinality at all; rather, each fuzzy set has n cardinalities, where n is the number of possible degrees of membership other than “definite non-membership.”)

In the absence of a theory of sets of indefinite cardinality, it might be better for the depraved semanticist investigating languages in which vague identity is expressible to try to get along without such *objects* as universes, pairings and frontiers. I am not sure whether this could be done. There are various devices for making statements about things collectively that do not require quantification over collections of things. Languages embodying such devices fall far short of the expressive power of set theory, but they may contain sufficient resources for talking about things collectively to meet the needs of the depraved semanticist investigating vague identity. The devices I have in mind are plural referring expressions (‘Tom and his colleagues’), predicates of variable polyadicity (‘are having a picnic’), plural variables (‘the x s’, ‘the y s’), and plural quantifiers (for any x s’, ‘for some y s’). I have not tried to construct a semantics for vague identity using only these resources for collective reference. It would be interesting to see whether this was possible.

Objection Four

“Various *truths of reason* fail in your system. That is, your semantics fails to confer validity on inference-forms that have always been accepted. We have already seen that the Law of the Excluded Middle fails on the only plausible extension of your system to include disjunction. The same point could be made about the Law of Non-Contradiction (given the obvious definition of conjunction: a conjunction has the least value had by any of its conjuncts). Even if we stick with what is strictly speaking expressible in your sparse language, the Principle of the Transitivity of Identity fails. That is, the model

$$\{A, B, C\}, \{\{A, B\}, \{B, C\}\}, 'a' \text{ ref } A, 'b' \text{ ref } B, 'c' \text{ ref } C$$

is a counterexample to

$$a = b, b = c \mid - a = c;$$

it assigns $\frac{1}{2}$ to the premises and 0 to the conclusion. Moreover, the Principle of the Non-Identity of Discernibles fails in your system—or, rather, since the general principle is not expressible in your sparse language, instances of it do. For example, as you have noted, I_4 is invalid. The model

$$\{A, C\}, \{\{A, C\}\}, 'a' \text{ ref } A, 'b' \text{ ref } A$$

assigns $\frac{1}{2}$ to its premises and 0 to its conclusion.”

Reply

I have already remarked on the Law of the Excluded Middle. I think that the validity of this “law” depends on the assumption that every proposition is definitely true or definitely false. But if anyone *must* have it, he can secure it by the “supervalue” trick outlined above. A similar remark applies to the Law of Non-Contradiction. I think that the validity of this law, too, depends on the assumption that every proposition is definitely true or definitely false. (Suppose it is neither definitely true nor definitely false that John is rich. Suppose Tim says, “John is rich” and Tom says, “John is not rich.” Is it definitely false that they’re both right? Remember, it’s definitely true that they’re both *sort of* right.) Again, however, anyone who insists on having this “law” can have it by employing the “supervalue” trick. In any case, since Evans’s argument employs no binary connectives, *no* way of assigning values to sentences involving binary connectives will affect the implications of our semantical fragment for his argument.

As to the failure of transitivity of identity, isn't that just what we should want? Suppose that *Omega* enters the Cabinet, and that Aleph (who is only indefinitely identical with Omega) emerges, and that Aleph then re-enters the Cabinet, from which Beth subsequently emerges. Suppose that this process continues till Tav emerges from the Cabinet. *Must* we say that it is neither definitely true nor definitely false that Alpha is identical with Tav? Couldn't the twenty-four operations of the Cabinet have so diluted whatever factor is the ground of personal identity that Alpha and Tav are definitely *not* identical? Must *logic* rule this out? Must we allow logic to dictate to us that if all members of the chain: indef Alpha = Omega, indef Omega = Aleph, indef Aleph = Beth, ..., indef Sin = Shin, indef Shin = Tav are true, then it can't be that Alpha and Tav are definitely two distinct objects? If we accept the transitivity of identity as a logical truth, we are accepting this consequence as a decree of logic.¹²

Well, *isn't* the transitivity of identity a truth of reason? Can't my own words be quoted against me? (I once wrote, "Anyone who denies the transitivity of identity simply does not understand the difference between the number 1 and the number 2."¹³) No and no. What is a truth of reason—and what I was *calling* 'the transitivity of identity'—is the principle that should in the present context be called 'the transitivity of *definite* identity': if it is definitely true that $a=b$, and it is definitely true that $b=c$, then it is definitely true that $a=c$. And this principle *is* endorsed by our semantical fragment: any model that assigns 1 to ' $a=b$ ' and to ' $b=c$ ' will also assign 1 to ' $a=c$ '. We may note that this fact constrains anyone who accepts our fragment in the following important way: if he accepts ' $a=b$ ' and ' $b=c$ ', he must also accept ' $a=c$ '.

But suppose someone insists that identity have *this* feature: if it is indefinite whether $a=b$ and indefinite whether $b=c$, then it cannot be definitely false that $a=c$; we can give this person what he wants easily enough. We simply impose the following condition on pairing: For any model, if three objects x , y , and z belong to the universe of that model, and if that model pairs x with y and y with z , it must also pair x with z . We may note that the model M , our counterexample to I_3 , satisfies this condition. Therefore, adopting this restriction on pairing would not affect our discussion of Evans's argument.¹⁴

Let us finally examine the invalidity of the Principle of the Non-identity of Discernibles. I will make three points.

(i) What we normally call ‘the Non-identity of Discernibles’ certainly has the aspect of a truth of reason. But perhaps what we normally think of under the description ‘the Non-identity of Discernibles’ is a principle that would be better described in the present context as ‘the *definite* non-identity of *definite* discernibles’. And this principle is valid according to the semantical fragment I have proposed. (Cf. our remarks about the transitivity of *definite* identity.) Consider any inference-form

$$a \text{ has } F, \sim b \text{ has } F \mid - \sim a = b,$$

where F is an abstract. Any model that assigns 1 to both premises must include the referent of ‘ a ’ in the extension of F and exclude the referent of ‘ b ’ from both the extension and the frontier of F . Thus, the referents of the two constants in that model must be distinct and not paired, and the model will assign 1 to the conclusion.

(ii) Is it so very implausible that (if there is such a thing as the indefinite possession of a property at all) the Principle of the Non-Identity of Discernibles can take us from $\frac{1}{2}$ to 0? Suppose it is not definitely true and not definitely false that x has the property F . Then the proposition that x has F and its denial will both have the value $\frac{1}{2}$; but the proposition that x is not identical with x should, it seems, nevertheless receive the value 0. Consider a formal instance of the Non-identity of Discernibles that contains only *one* constant, ‘ a ’; let the abstract it contains be that of I_4 :

$$a \text{ has } \hat{x} \text{ indef } x = a, \sim a \text{ has } \hat{x} \text{ indef } x = a \mid - \sim a = a.$$

Any model that pairs the referent of ‘ a ’ with something will assign $\frac{1}{2}$ to each of the premises of this inference-form, and any model at all will assign 0 to its conclusion. Perhaps it is implausible to assign $\frac{1}{2}$ to either of these premises. We have already considered that question. Given that we do that, however, we should certainly want the conclusion to have the value 0. And should matters really be different if we assigned a second constant ‘ b ’ the same referent as ‘ a ’ and replaced the first occurrence of ‘ a ’ in the second premise and one of the occurrences of ‘ a ’ in the conclusion with ‘ b ’?

It is important to note that the failure of the Non-identity of Discernibles in our semantical fragment does not turn on our ‘‘controversial’’ rule for assigning values to ascription sentences. It will arise in any semantics that allows a sentence of the form ‘ a has F ’ to take the value $\frac{1}{2}$. And that, surely, must be a feature of any plausible semantics that applies to a language that attempts to represent the possession of properties that have

borderline cases. Suppose, for example, that a certain man is a borderline case of a wise man, and that he bears the names ‘Tom’ and ‘Tim’. It is hard to see how to avoid the conclusion that both of the following arguments have “sort of true” premises and a definitely false conclusion:

Tom has wisdom, \sim Tom has wisdom $\mid - \sim$ Tom = Tom

Tom has wisdom, \sim Tim has wisdom $\mid - \sim$ Tom = Tim.

(iii) Our adherence to the Principle of the Non-Identity of Discernibles is based on our adherence to an even more fundamental truth of reason, Leibniz’s Law or the Principle of the Indiscernibility of Identicals,¹⁵ which we may represent schematically as follows:

$$a = b, a \text{ has } F \mid - b \text{ has } F.$$

The counterexamples to the Non-Identity of Discernibles that assign $\frac{1}{2}$ to the premises of that inference-form and 0 to its conclusion will not be counterexamples to Leibniz’s Law: they will assign 1 to the identity premise, $\frac{1}{2}$ to the other premise, and $\frac{1}{2}$ to the conclusion. (In two-valued logic the two principles are equivalent; in three-valued logic—as the present case shows—it does not follow from the fact that $p, q \mid - r$ is valid that $p, \sim r, \mid - \sim q$ is valid.)

Nevertheless, Leibniz’s Law is not valid in our semantical fragment. It would become valid, however, if we adopted the condition on pairing that secures the transitivity of identity. (The proof is left as an exercise.) Therefore, any philosopher who wishes to have Leibniz’s Law can have it at the same price as the Transitivity of Identity.

This concludes my examination of possible philosophical objections to the semantical fragment I have presented. I know of no others. While I should not want to say that these objections, taken collectively, have no force whatever, they seem to me to be very far indeed from being decisive. I conclude that the friends of vague identity can take Evans’s argument in their stride.

NOTES

* I have greatly benefitted from correspondence and conversation with, and access to the unpublished work of, a great many people. I am grateful to George Bealer, Monte Cook, Christopher Hill, Frances Howard, Hans Kamp, Ralph Kennedy, David Lewis, Alastair Norcross, Terence Parsons, Jeffrey Pelletier, Alvin Plantinga, David Sanford, Robert Stalnaker, and Richard Taylor, each of whom provided me with one or more of these benefits.

I wish especially to thank Nathan Salmon for his two extensive sets of comments on a draft.

Parts of drafts of the paper were read at a symposium on "Objects" at the 1986 meetings of the Central Division of the American Philosophical Association, at the Fall 1987 meeting of the Creighton Club, and to colloquia at the University of Notre Dame, Memphis State University, Vanderbilt University, and California State University, Northridge. Various members of the audiences on these occasions raised questions that have added to, subtracted from, or modified the paper. I have tried to remember these people in the above list.

1. Why does the example involve the operations of a bizarre science-fictional contraption and why does it involve such complicated objects as human beings? Could there not be an at least equally effective example that involved only the ordinary mutations of the world and puddles (the puddle I stepped in yesterday and the puddle I stepped in today) or piles of trash (the pile I swerved to avoid last week and the pile I swerved to avoid today) or clouds (the cloud I was looking at ten minutes ago and the cloud I am looking at now)? Well, yes there could—provided that (in the strict, philosophical sense) there *are* such things as puddles and piles and clouds; and provided that puddles and piles and clouds can at least to some small extent survive (in the strict, philosophical sense of 'survive') the ordinary mutations of the world. But some might deny that, e.g., the water molecules within a water-filled depression really do add up to or compose an object; some might say, that is, that there is only a convention according to which we sometimes talk as if they did. (This is, in fact, a thesis that I myself find very attractive.) Others might say that the molecules within the depression do indeed add up to an object, but they have always (whenever they all existed) added up to numerically the same object, even when they were (as they usually have been) scattered to the four corners of the earth. According to the former view, puddles do not provide examples of vague identity, for the very simple reason that there *are* no puddles. According to the latter view, puddles do not provide examples of vague identity, for it is only by convention that we treat the molecules that are in a certain depression at two different times as constituting the same object; in sober truth—this view tells us—no one ever steps twice into the same (object of the sort that can momentarily be a) puddle. But many of us would be unwilling to take a similar line about *people*. Many of us want to say that people really do exist, and that they strictly persist through the ordinary mutations of the world (which are mutations that they are really "directly" subject to, and not ones that they are merely subject to vicariously—as a Cartesian ego is subject to changes in the material world only in the sense that it is bound to a human body that "directly" undergoes those changes). For many of the readers of this paper, the (strict) existence and the (strict) persistence of (material) human persons will be a "non-negotiable" matter; whereas the strict existence and persistence of puddles, piles, and clouds will be eminently negotiable: they will be things we should abandon with a light heart and an easy ontological conscience the moment we became convinced that belief in them committed its adherents to belief in vague identity.

But, of course, anyone who does believe that there really *are* puddles and that one can really step twice into the *same* puddle should feel free to construct an example involving puddles and to substitute it in his mind for the "Cabinet" example.

A second question that may puzzle some people. Why not regard persons (and other persisting objects) as being four-dimensional objects, "constructs" out of momentary person-stages that are bound together into a four-dimensional whole by certain unifying relations of causation and spatio-temporal continuity? (For it is fairly easily seen that on that view of persons—and other persisting objects—, any vagueness that infects statements about "identity across time" can plausibly be said to have its source in vague

general terms.) The short answer is: I do not accept this picture of persisting objects. I believe that one and the same, literally one and the same, three-dimensional object can exist at two different times. But I do not propose to defend this position in the present paper. Adherents of "four-dimensionalism" will find that the present paper is addressed to their condition only to the extent that it displays the difficulties that must be surmounted by views alternative to their own. (We shall briefly examine some alleged examples of vague identity that do not involve identity across time; but I do not think that these examples are convincing enough "on their own" to support a belief in vague identity. Doubtless the four-dimensionalists, lacking the independent motivation for believing in vague identity that is provided by cases involving identity across time, will prefer to deal with these examples by some *ad hoc* device that operates within the bounds of standard logic.)

2. "Can There Be Vague Objects?," *Analysis* 38.4, Oct. 1978, p. 208.
3. At this point in a draft of the paper I made the parenthetical remark, 'Nathan Salmon has a different argument—but I think it is equivalent—for the same conclusion'. Correspondence with Salmon has convinced me, however, that the question whether Salmon's argument—see his *Reference and Essence* (Princeton: Princeton University Press, 1981), pp. 243-45—and Evans's argument are in any interesting sense equivalent is too delicate to be relegated to a parenthetical remark. For all I say in this paper, therefore, it may be that Salmon's argument is a cogent refutation of the coherency of vague identity. I hope to discuss Salmon's argument elsewhere.
4. *Op. cit.* Evans's puzzling words make sense if we postulate that he supposes that the dual of 'indef' is 'it is definitely true that'. But this, of course, is false: the dual of 'indef' is 'it is either definitely true or definitely false that'. How to expand a valid deduction of ' $\sim\alpha = \omega$ ' from 'indef $\alpha = \omega$ ' into a valid deduction of ' \sim indef $\alpha = \omega$ ' from that premise is by no means the trivial problem it might at first glance appear to be. In an unpublished paper, "The Good, The Bad, and The Ugly," F.J. Pelletier presents a proof of ' \sim indef $\alpha = \omega$ ' from 'indef $\alpha = \omega$ ', a proof that (essentially) contains Evans's deduction as a part. But this proof depends on some strong second-order principles, such as 'indef $AxFx \cdot \leftrightarrow \cdot \exists x(\text{indef } Fx) \& \sim \exists x \text{ deff } Fx$ '. (Here 'deff' means 'it is definitely false that'.) In this proof, 'F' is not a predicate-letter but a true variable (both 'A F' and 'E F' occur in the proof).
5. In writing names of models in this form, we adopt the convention that 'A', 'B', 'C', and so on, always have distinct referents when they are used within the name of a given model.
6. In correspondence, Nathan Salmon has raised an interesting and subtle objection to the "ascription" rule:

The rule you propose is the correct valuation rule not for the ascription ' α has $\hat{x}\phi$ ', but for the more complex assertion 'Something identical with α has $\hat{x}\phi$ '. Of course, classically these are equivalent. But once identity is claimed to be vague, these two assertions come apart. . . . Intuitively, the first assertion is false whenever . . . the referent of α belongs neither to the extension nor the frontier of ' $\hat{x}\phi$ '. . . . By contrast, the second assertion will be neither true nor false if the referent of α [belongs neither to the extension nor the frontier of ' $\hat{x}\phi$ '] but in addition . . . some "fringe-referent" of α ends up in either the extension or the frontier of ' $\hat{x}\phi$ '. . . .

It seems to me that the persuasive power of this objection (which I concede) can be undermined if we think in terms of definite descriptions—if it is appropriate to call them "definite" in the present context. Suppose, as seems plausible, that we regard ' α ' as having (whether or *not* identity as vague) the same semantical properties as 'the x such that $x = \alpha$ '. Making the appropriate substitution in Salmon's objection, we have

Intuitively, 'the x such that $x = \alpha$ has $\hat{x}\phi$ ' is [determinately] false whenever the referent of 'the x such that $x = \alpha$ ' belongs neither to the extension nor the frontier of ' $\hat{x}\phi$ '.

Is this intuitive? Is it in fact plausible in the case in which the identity-predicate that occurs in the description has itself a non-empty frontier? Let us generalize this question. How, in general, should we evaluate sentences of the form 'the x such that Gx has $\hat{x}Fx$ ' when G is vague? The interesting case is this: there is exactly one definite case of G and there are one or more borderline cases of G 's. Consider, for example, the sentence 'the x such that x is tall has wisdom'. Suppose the universe consists of two objects, one definitely tall and not definitely wise—in fact, let's say, definitely foolish—, and the other borderline-tall and, let's say, definitely wise. If we have in our semantical apparatus only one alternative to definite truth and definite falsity, isn't it highly plausible to suppose that we should assign that one alternative to 'the x such that x is tall has wisdom'? (If there were more than one alternative to definite truth and definite falsity—in a four-or-more-valued system—, we might want to consider evaluating this sentence differently in the case described and in the case in which one thing is definitely tall and that thing is borderline-wise and nothing is borderline-tall.) Should we not say the same thing about 'the x such that $x = \alpha$ has wisdom' that we have said about 'the x such that x is tall has wisdom'? The predicate ' $x = \alpha$ ' and the predicate ' x is tall' would seem (on the assumption there are vague identities) to differ in only one respect that is of present interest: ' $x = \alpha$ ' must have exactly one item in its extension (obviously an important feature when the topic is definite descriptions), whereas ' x is tall' may have any number of items in its extension, including 0. But shouldn't we treat 'the x such that $x = \alpha$ has wisdom' exactly as we treat 'the x such that x is tall has wisdom' in those cases in which ' x is tall' has exactly one item in its extension? And shouldn't we treat ' α ' just as we treat 'the x such that $x = \alpha$ '? That, at any rate, is how I *have* treated ' α ': the referent of ' α ' is the member of the extension of ' $x = \alpha$ ', and the fringe of ' α ' is the frontier of ' $x = \alpha$ '.

It may be that there is something to Salmon's contention that "once identity is claimed to be vague, the two assertions ' α has $\hat{x}\phi$ ' and 'Something identical with α has $\hat{x}\phi$ ' come apart." If so, I would suggest, this "coming apart" cannot be represented in a three-valued semantics. The coming apart should, if it exists, be represented in a four-or-more-valued semantics by the two sentences' in some cases receiving different "intermediate" values—and never by one of them receiving one of the intermediate values and the other definite truth or definite falsity. (And, of course, in a four-or-more valued semantics, $\text{indef } p$ should receive the value 1 if p takes on *any* of the intermediate values, and 0 otherwise.)

We should note that we can construct a semantics that accommodates the scruples of those who insist on the rule ' a has $\hat{x}Fx$ has the value $\frac{1}{2}$ if *and only if* the referent of a belongs to the frontier of $\hat{x}Fx$ ', and which nevertheless fails to confer validity on Evans's reasoning: we simply insist that the frontier of a predicate include anything indefinitely identical with anything in the predicate's extension or frontier (provided that it is not "already" in the predicate's extension). We may do this as follows. We have specified (relative to a model) an extension and frontier for each predicate. Replace the word 'frontier' in this specification at each of its occurrences with 'proto-frontier'. Now define the *frontier* assigned to a predicate P by a model m as the union of the proto-frontier that m assigns to P with the class of all things that (a) do not belong to the extension of P and (b) are paired with members of P 's extension or P 's proto-frontier. And let all be as before, except that ' a has $\hat{x}Fx$ ' is to receive the value $\frac{1}{2}$ if and only if the referent of a belongs to (what is now called) the frontier of $\hat{x}Fx$. I think that this semantics is

equivalent to the one I have proposed, although I have not carefully investigated this. At any rate, it certainly has the consequence that M is a counterexample to I_3 .

I should finally point out that the second of Salmon's sentences actually contains a possible "source" of vagueness that is certainly not to be found in the first: the existential quantifier. (I said at the beginning of this paper that I thought that there were several ways in which vagueness could exist "in the world"; I will now briefly touch on a "way" that does not involve identity.) In "When Are Objects Parts?" (*Philosophical Perspectives* I, 1987; pp. 21-47) I have provided some motivation—the idea derives from a suggestion of David Lewis's—for thinking that sentences of the form ' $\exists xFx$ ' should sometimes receive the value $\frac{1}{2}$ even though the frontier of F is empty. (See pp. 43-5.) I note in passing that if this is indeed true, then the principle cited in n. 4 that Pelletier uses in his deduction of a contradiction from 'indef $a = b$ ' has counterexamples.

7. "But shouldn't we conclude, by parity of reasoning, that it is not definitely false that it is indefinite whether Omega is identical with Omega—and, more generally, that I_1 is invalid? Omega is, after all, *sort of* identical with Alpha, and it is indefinite whether Alpha is identical with Omega." But this is semantically impossible for reasons that have nothing to do with identity: it follows from the value-table for 'indef' that any formula of the form 'indef indef p ' must receive the value 0. "And a very plausible table it is. But maybe your attempted justification of the 'ascription' rule is inconsistent with the very plausible conception of indefiniteness embodied in that table. If a piece of reasoning exactly parallel to your reasoning has the consequence that 'indef $\omega = \omega$ ' should receive the value $\frac{1}{2}$ (a value no formula of the form 'indef p ' can have)—well, so much the worse for your reasoning. Consider the 'suggested strategy', which has guided your reasoning. Let us apply that strategy "directly." Omega is *sort of* identical with something (Alpha) that is indefinitely identical with Omega. Isn't that at least a *weak* reason for calling 'indef $\omega = \omega$ ' true? And doesn't the proposed strategy therefore tell us to assign $\frac{1}{2}$ to 'indef $\omega = \omega$ '? And isn't that strategy therefore defective?" It seems to me, on reflection, that there could not be a reason, even a very weak reason, for hesitating over whether to call a sentence that consists of '=' flanked by two occurrences of the same symbol true. But I find that I do not have the same sort of strong conviction about the outcome of inserting two occurrences of the same symbol in the schema ' \dots has $\exists x = \dots$ '. *Maybe* there could be at least a weak reason for hesitating to call such a sentence true. (Of course, if someone *did* think that there could be at least a weak reason for hesitating about whether to call a sentence of the form ' $a = a$ ' true, and if he accepted the proposed strategy, he would probably wish to construct a semantics according to which I_1 was invalid. Presumably his semantics would assign 1 to an identity sentence only if the referent of its term(s) were not paired with anything.) This pair of convictions is possible for someone only if he lacks an overriding confidence in the validity of property abstraction. I do. Russell's Paradox, as I have remarked, shows that property abstraction *can* fail. But if it can fail in one sort of case, can we be perfectly confident that that is the only sort of case in which it can fail?
8. They will not, if they are wise, claim even to have presented an *argument* for the conclusion that Evans's reasoning is invalid—or not simply in virtue of having constructed a formal semantics according to which that reasoning is invalid. But here is an outline of an argument for that conclusion: (1) The "Cabinet" case (or perhaps a case involving puddles and piles of trash) makes it plausible to suppose, antecedently to any argument, that there do exist cases of vague identity, and hence that any argument for the non-existence of such cases contains an error. (2) The semantical fragment presented herein puts forward a reasonably plausible candidate for the office "error in Evans's reasoning"—at least (I predict) it will be regarded as a reasonable candidate by someone who

antecedently believes that there probably is such an error; I do not predict that it will be regarded as a plausible candidate by those who accept Evans's conclusion. This, of course, is only an argument. It is not a proof. There are, I think, few if any proofs of interesting philosophical theses. The argument could be replied to, like (almost?) all philosophical arguments, and the reply could be replied to, and so on, *ad infinitum*. What someone said of the study of history can with at least equal justice be said of philosophy: philosophy is argument without end. But I don't think that the above argument is a bad argument—as philosophical arguments go. A philosophical argument does not count as a failure if it does not force assent to its conclusion. It is enough that it lend some support to that conclusion.

9. "Supervalues" are the invention of Bas van Fraassen. See "Presupposition, Implication, and Self-Reference," *Journal of Philosophy*, 1968, pp. 136-52.
10. More exactly: to no philosophical need having to do with vagueness. Supervalues may have their place if one is evaluating sentences in a language having "truth-value gaps" (a language in which no value at all is assigned to some sentences), or if the intended interpretation of the "intermediate" value has some other basis than vagueness. (Consider for example, the semantical problem facing an Aristotelian who believes that the sentence 'There will either be a sea battle tomorrow or there will not' is true and who also believes that both of its disjuncts have the intermediate value.)
11. The following "Reply" contains the germ of what I would say in response to Salmon's argument against the possibility of vague identity, but I do not claim that it will do as it stands as a criticism of his argument.
12. Frances Howard has pointed out to me the philosophical utility of allowing ' $a=c$ ' to have the value 0 when ' $a=b$ ' and ' $b=c$ ' have the value $\frac{1}{2}$. This utility is well illustrated by the problem of the Ship of Theseus. Let us borrow three useful terms of Jonathan Bennett's: *the Original Ship* (O), *the Reconstructed Ship* (R), and *the Continuous Ship* (C). If we could get away with it, it would be pleasant to say that ' $C=O$ ' and ' $O=R$ ' were both *sort of* true, and that, nevertheless, ' $C=R$ ' was definitely false. It would be a shame if logic (as opposed to metaphysical considerations about artifacts) forbade us that pleasant option.
13. "The Doctrine of Arbitrary Undetached Parts," *Pacific Philosophical Quarterly* 62 (1981) pp. 123-37. See p. 128.
14. Doesn't rejecting the transitivity of identity violate my "proposed strategy" of assigning $\frac{1}{2}$ to a sentence if there is the least excuse for doing so? If ' $a=b$ ' and ' $b=c$ ' both have the value $\frac{1}{2}$, doesn't their both having that value constitute a good reason for not assigning 0 to ' $a=c$ '? This question misconstrues the strategy I have proposed. The misconstrual is facilitated by my occasional use of the graphic but imprecise wording "assign the value $\frac{1}{2}$ if there is the least excuse for doing so." Here is a statement of the strategy that is less open to misconstrual: *Given* the information contained in a model (given certain objects, given a specification of which objects are paired, and given an assignment of referents to constants drawn from among those objects), assign the value $\frac{1}{2}$ to a sentence on that model if there is the least excuse for doing so. The strategy, in other words, constrains the use made of the information contained in a model, and does not constrain what information may be contained in a model; if a model *does* contain the information that ' a ' and ' c ' denote distinct objects that are not paired, there can be no excuse for assigning any value but 0 to ' $a=c$ ', no matter what other information may be contained in the model. One could, of course, adopt an *extension* of the "proposed strategy," one that constrained the information contained in a model in the following way: if *some* of the information contained in a model provides the least excuse for assigning $\frac{1}{2}$ to a sentence, then the rest of the model must be "filled in" in such a way

that $\frac{1}{2}$ is assigned to that sentence by the model. This extended strategy would, I suppose, lead us to adopt the constraint on the pairings included in a model that is mentioned in the paragraph to which this note is attached.

15. This is an adaptation of a point made by Terence Parsons in his important paper, "Entities without Identity." (*Philosophical Perspectives* 1, 1987, 1-19). I say "an adaptation" because Parsons's point is made in relation to "the contrapositive of Leibniz's Law" and "Leibniz's Law," which correspond to, but are not the same as, what I call "the Principle of the Non-Identity of Discernibles" and "Leibniz's Law." What Parsons calls "Leibniz's Law" is a more general principle than the one I call by that name.

Let me take this opportunity to commend Parsons's paper, which contains a different approach to the problem of vague identity from mine, one that in many respects I like better than my own. Parsons's approach has the consequence that the property abstracts in Evans's argument fail to denote anything. And this is not the result of an *ad hoc* stipulation, but is extremely well-motivated.