



## Australasian Journal of Philosophy

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/rajp20>

### On Gödel's ontological proof

A.P. Hazen<sup>a</sup>

<sup>a</sup> University of Melbourne

Published online: 28 Jul 2006.

To cite this article: A.P. Hazen (1998) On Gödel's ontological proof, Australasian Journal of Philosophy, 76:3, 361-377, DOI: [10.1080/00048409812348501](https://doi.org/10.1080/00048409812348501)

To link to this article: <http://dx.doi.org/10.1080/00048409812348501>

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access

and use can be found at <http://www.tandfonline.com/page/terms-and-conditions>

## ON GÖDEL'S ONTOLOGICAL PROOF

A.P. Hazen

### I. Logical Exordium

A frequent theme in 20th century philosophical reflection on modality has been the feeling that *de re* modality and 'quantifying in' are somehow more problematic than mere *de dicto* modality. Perhaps the most influential proponent of this view was W.V. Quine, who in a series of papers starting in the 1940's argued that *de re* modal assertions were uninterpretable. The background of his argument was a general philosophical view of modality, deriving both from Carnap and from C.I. Lewis, that identified *necessity* with *analyticity*. Now, famously, Quine had his doubts about the notion of analyticity, but he was willing to consider it hypothetically: if analyticity was a coherent notion, and if one defined necessity as analyticity, the application of modal operators to complete sentences—modality *de dicto*—could be understood, but, he argued, this was insufficient for the interpretation of *de re* modal locutions, which required some further notion of an attribute holding *essentially* of an object. Following the development of formal, model-theoretic, semantics for quantified modal logic in the 1960s (which provided the essential framework for the more metaphysical discussions of modality in the 1970s), Quine's extreme view has become less popular than it once was. Still, although many philosophers recognize a notion of essence allowing for the interpretation of *de re* modalities as coherent, or even as deeply embedded in our conceptual scheme, there is a lingering feeling that this notion is less basic and less robust than the notions on which the interpretation of *de dicto* modalities rest: that it depends on some *further* features of our conceptions of objects, features that are in some way at least analogous to linguistic conventions (cf. Sidelle [24]).

The implication of this feeling for logic is that the *de dicto* portions of (quantified) modal language and logic are better understood than and more secure than—less likely to be undermined by changes in our conceptions of the world—than the *de re*. A *de dicto* fragment of modal logic, then, might be seen as a particularly sound and perspicuous framework for philosophical argument, and an analysis of just which portions of a contentious philosophical argument can be accommodated within it, and (if an argument cannot be rendered entirely *de dicto*) just where and how it oversteps these bounds, can provide useful insight. Certain relevant points of formal modal logic have been clear, now, for over twenty years. Standard systems of First-Order modal logic with Identity are not committed to any substantive essentialism: by the results of Parsons [22], any such system can be extended consistently (and without collapsing modal distinctions) to one with strong 'anti-essentialism' principles, in effect making sentences with *de re* modalities into redundant *façons de parler* equivalent to certain other, purely *de dicto*, sentences. Parsons proves his result for systems containing the Barcan Formula and its converse, and so having semantic interpretations in which the domain of individuals is the same for all

possible worlds, but with minor modifications and additions (such as the adoption of something like the 'falsehood convention' for atomic formulas at worlds where not all of their terms denote) his methods can be adapted to prove a similar result for systems without these axioms, systems that allow the domain of individuals to differ (and differ in cardinality) from world to world (Hazen [13]<sup>1</sup>). Moreover, these results do not simply hold for the pure logics, but extend in a strong and interesting way to arbitrary modal theories. Assume a *theory* (set of sentences closed under logical consequence) formulated in the language of First-Order modal logic and consistent (either in the sense of the logic with or the logic without the Barcan Formula and its converse) by the lights of standard First-Order modal logic with identity. Given an interpretation of the language—say one in which the quantified variables are thought of as ranging over all non-abstract entities, the predicates as expressing particular empirical properties and relations, and the modal operators as expressing 'logical' or 'conceptual' or 'metaphysical' necessity and possibility—the set of all *true* sentences of the language would be such a theory. Then the subset of the theory consisting of all its purely *de dicto* sentences—sentences in which no proper names occur in the scope of modal operators and in which no quantifier outside the scope of a modal operator binds occurrences of a variable inside the scope of that operator—is consistent by the lights of the logic supplemented by the 'anti-essentialism' principles mentioned above: otherwise put, there is a theory with exactly the same purely *de dicto* theorems as the one we started with, but according to which all *de re* modality is trivial.

The discussion in the preceding paragraph is not, of course, meant to suggest that the judgments expressed in sentences containing *de re* modal locutions are unimportant, or that we could trivially abandon our essentialistic opinions in order to adopt an anti-essentialist modal logic. On the contrary: it might well be argued that to adopt anti-essentialism principles as modal axioms would be to renounce key features of our ordinary ontology of individual objects, and to turn the values of the variables of our quantified language into mere 'indices,' thereby making the language into something like a 'feature-placing' (Strawson [26]) or 'coordinate' (Carnap [6]) language. It is meant to suggest, rather, that purely *de dicto* sentences express a relatively stable subject matter, one that is to a considerable degree separable from particular commitments about the essences of things, and that the usual modal systems respect this separation. In particular, it suggests that these systems provide a suitable framework in which to study the logical relations between such sentences, one which does not 'entangle' them with questions about essence. We may, then, assume as the First-Order part of our logic in what follows, the principles formalized in such a system. For definiteness: S5 (since there is widespread agreement that this is the correct propositional modal logic for the logical or metaphysical modalities), without the Barcan Formula and its converse (since this version seems to prejudge the least about what exists necessarily, and makes plausible the interpretation of the existential quantifier as expressing *existence*), but with an unrestricted rule of substitution of identicals (since, on the one hand, Kripke [15] has argued for its validity and, on the other, it is compatible with strong anti-essentialism principles).

<sup>1</sup> For those who prefer published sources, cf. Fine [9] and the discussion in Kripke [17].

Reference, in modal contexts, to *properties* is usually (though with exceptions: cf. [27]) thought of as not raising questions of essentialism. Occurrences of predicates, unlike occurrences of proper names, in the scope of a modal operator are not thought of as rendering a sentence *de re*. Moreover, a typical instance of a *comprehension principle* for modal Second-Order logic would say that, for a given formula containing a free individual variable, *there is* a property, *F*, such that necessarily (that is, in any arbitrary possible world) *F* holds of all and only those individuals which (in that world) satisfy the formula. To say this, however, is to quantify into the scope of the necessity operator: Second-Order modal logic, it seems, requires us to be essentialists about properties, even if we reject essentialism about individuals. It is possible to be bothered by this disparity, but it seems to me that it is also possible to regard it as a natural concomitant of the different roles properties and individuals play in our thinking. Modal operators seem formally to amount to (notationally simplified) quantifiers over possible worlds, and possible worlds—‘ways things could have been’—have often been explained as special cases of propositions (‘complete novels’) or of properties (maximally strong contingent properties of spatio-temporally unbounded fusions of individuals), so quantifiers over these conceptual entities should be expected to interact more freely with modal operators than do quantifiers over individuals.

As a logical framework for arguments about properties, then, we may add Second-Order variables to our First-Order language, and postulate the usual modal quantificational logic as holding for them as well as for individual variables. For our Second-Order quantification, however, it is reasonable to postulate, in addition to the principles imposed on First-Order quantification in modal logic, the Barcan Formula and its converse: this simplifies matters, and, though the assumption that properties all exist necessarily is open to contention (cf. the discussion of ‘alien universals’ in [20] and [14]), doubts about it do not seem relevant to the topic at hand. We will also need comprehension principles, and this raises delicate issues. The simplest and most obvious principles, which seem to have been assumed by Gödel, Scott, and their commentators, would be to postulate an axiom scheme along the lines of

$$\Box \exists F \Box \forall x (F[x] \leftrightarrow \dots x \dots),$$

where  $\dots x \dots$  is a formula not containing the (monadic) Second-Order variable *F* and any free variables other than *x* in  $\dots x \dots$  are bound by universal quantifiers at the extreme left of the sentence (and, not that we will make any use of it here, similarly for dyadic and higher-degree relations): in words, that necessarily (since it is an axiom!) there is a property of which it is necessary that it holds of all and only the individuals satisfying the formula. (These principles are enough to support Gödel’s argument, but do not by themselves suffice for a satisfactory general modal property theory.) If identity, for properties, is defined as necessary coextensiveness, the analogues of the usual principles of the logic of identity will be provable.

The complication comes from a consideration of what are variously called *relational*, or *impure relational*, or *haecceitistic* properties: properties defined by reference to a specific individual, such as *having studied under Alfred Tarski*, or, in the extreme case, such haecceities as *being identical to Dana Scott*. There is some dispute among metaphysicians as to whether such ‘properties’ are genuine properties at all—they would

certainly not be considered universals by a theorist of Armstrong's 'sparse' school (cf. Lewis [19])—and even if they are recognized as properties, some—Plantinga calls them 'existentialists'—who are happy to admit the necessary existence of *qualitative* or *characterizing* properties would hold that haecceities exist only if the individuals whose haecceities they are exist. We do not need to commit ourselves on these questions. What is important is to note that unrestricted reference to relational properties allows a trivial reduction of *de re* assertions about individuals to *de dicto* ones. Instead of asserting, *de re*, that Scott might have written best-selling historical novels, we may assert that his haecceity might have been coinstantiated with the property of writing best-selling historical novels. Obviously, a sceptic about the meaningfulness of *de re* modal locutions would also be sceptical of the genuineness, or at least of the necessary existence, of haecceitistic properties. It will be useful, therefore, in examining the presuppositions of (variants of) Gödel's argument, to consider a variant theory of properties, on which property variables are restricted to range only over non-relational or non-haecceitistic properties. Such a theory would add the further restriction on the comprehension schema that  $x$  should be the only individual variable occurring free in  $\dots x \dots$ , and that no proper name (individual constant) occurs in that formula.

If we identify necessarily coextensive properties, as we will in what follows, properties—on either conception—form a Boolean algebra. Since no stricter sense of entailment (such as that studied in Anderson and Belnap [3]) is at issue, we will, for brevity, follow the sloppy usage of some commentators on Gödel and say that one property *entails* a second if the first strictly implies the second: if, that is, it is necessarily the case that any individual with the first will have the second as well. A (or, on the mentioned identification, the) contradictory property entails all properties, and necessary co-extensiveness is the same as mutual entailment.

## II. Gödel's Argument

Gödel, in a cryptic note [11] given to Dana Scott in 1970, introduces the notion of a *positive* property. (Thus the formalized version uses a Third-Order constant, the predicate  $P(F)$ , expressing the positivity of the property  $F$ .) The plausibility of his 'axioms' and the theological relevance of its conclusion depend on the interpretation of this notion; Gödel says that it means 'positive in the moral aesthetic sense (independently of the accidental structure of the world)', but also allows that it may mean 'pure "attribution" as opposed to "privation".' The evidence available from his notebooks suggests that he never found an interpretation of this notion that fully satisfied him, and it is perhaps best to assume that he thought of his ontological argument not as a conclusive proof of the existence of God, but as an attempt at a reconstruction of Leibniz's argument. In any event, he laid down certain axioms concerning the notion. Any property entailed by a positive property is positive, and the conjunction of two positive properties (the property, that is, that an individual has if and only if it has both of the given properties) is positive. Together, these amount to saying that the positive properties form a *filter* on the Boolean algebra of properties. (Gödel adds in a footnote that the positivity of conjunctions of positive properties holds for arbitrary numbers of conjuncts, not just for two: as we shall see, this includes infinite numbers.) Further, positivity is a non-contingent feature of a property: any property which is positive is necessarily positive, and no property which is not positive could be.

Another axiom said that, of any pair of properties consisting of a property and its negation (the property necessarily holding of all and only the individuals *not* possessing the first property), *precisely one* is positive. (Positive properties form an ultrafilter.) We are now in a position to prove our first theorem: positive properties are at least possibly instantiated. Proof: the contradictory property (self-non-identity) entails all properties, including its own negation (self-identity), so if it were positive both properties of such a pair would be positive. (As has been noted by commentators, the proof uses only part of the strength of the latest axiom: of a pair of a property and its negation *at most one* is positive.)

Gödel's remaining axiom is formulated in terms of a defined notion. First define a property to be an *essence* of an individual if it is a property, possessed by the individual (Gödel left this clause out in [11], but this appears to have been an oversight—it is included in related manuscripts), which entails every property the individual has. The terminology is somewhat unfortunate: the notion defined is close to Leibniz's notion of the complete concept of an object, but is not at all what current philosophical usage calls an essence. 'Essence' and 'essentialism' were bad words for most analytic philosophers of the 1960s—many found the doctrines of Kripke [16] downright shocking—but, with the flowering of modal metaphysics since, they have come to have well-understood and generally agreed meanings, defined in terms of *de re* modal locutions. An individual, *x*, has a property, *F*, *essentially* iff *x* has *F* and it is necessarily the case that *x*, provided it exists, has *F*. An *essence* of *x*, in this post-1970 sense, is a property which entails all and only the properties *x* has essentially. To avoid confusion, therefore, let us re-christen Gödel's notion: the property an individual has that entails every property it has is its *character*.

Character and essence (in the modern sense) are completely different. On the most usual assumptions, individuals do not have their characters essentially: Scott teaches at Carnegie–Mellon, so the property of teaching (in 1997) at Carnegie–Mellon is entailed by his character, but he *might* have stayed at Princeton (and thereby have had a different character). Essences, therefore, are logically weaker than characters: there are properties of Scott which are entailed by his character but not by his essence: his character itself, for one. (Leibniz seems to have thought otherwise. At least when speaking strictly, he held that, had the history of the world been different, not Scott but at best someone resembling Scott would have existed, or, as he put it in a passage quoted by Mates [21, p. 139]:

But, someone will object, whence comes it then that this man will assuredly do this sin? The reply is easy. It is that otherwise he would not be this man.

Leibniz, then, identified the notions of character and essence, and if we see Gödel as primarily concerned to reconstruct Leibniz's reasoning the terminology of his note is appropriate. For us non-Leibnitzians, however . . .)

Since, among the properties an individual has, there are the negations of all the properties the individual lacks, it follows from the definition that any property which is the character of an individual is an *atom* in the Boolean algebra of properties: a maximally strong possibly instantiated property, one which, for every property *F*, entails either *F* or the negation of *F*. Further, as Gödel notes, 'any two [characters] of *x* are necessarily equivalent,' or, as we may say given our identification of mutually entailing properties, an individual has at most one character. (Proof: if it had two, each would be one of the

properties entailed by the other.) Proving that (necessarily) every individual has at least one character is a bit trickier, even on the version of property theory that postulates relational properties and so allows free individual variables other than  $x$  in the formula  $\dots x \dots$  inserted in the comprehension scheme. The problem is that, given the placement of the modal operators in our modal comprehension axioms, if we tried the obvious instance of comprehension, we would be postulating the property, not of having the properties the given individual *actually* has—its character—but the property of having, at every world, the properties the given individual has *at that world*. Sobel [25] notes the trickiness, and takes as an additional axiom (in his derivation of some further theorems) the statement that every individual has a character: which is plausible enough, but not required for Gödel's own proof.

A general solution is perhaps best sought in a Third (or higher) Order logic having, in addition to the ordinary comprehension principles, analogues of the principles of 'Extensional Comprehension' discussed in Bressan [5] and Gallin [10]. A simpler solution, not requiring an extension of the language of modal Second-Order logic but sufficing to prove Sobel's special axiom and analogues of it mentioned below, would be to adopt an axiom scheme saying, for any condition we can write on a variable for properties, that there is a property which is the conjunction of all and only the properties satisfying the condition:

$$\exists H(\forall F(\Box \forall x(H[x] \rightarrow F[x]) \leftrightarrow (\dots F \dots))).$$

(Such a schema of infinitary conjunctions seems intuitively plausible; its analogue for non-modal Second-Order logic is derivable from comprehension.)

Next we define a property of individuals that Gödel calls, perhaps misleadingly, *necessary existence*: an individual has necessary existence (E, for short) just in case it has a necessarily instantiated character. (Gödel actually used a universal rather than an existential quantifier in his definition, which Sobel [25] noted had the harmless but aesthetically unpleasing consequence that any characterless individual would qualify as a necessary existent: in the case of individuals that have characters, the two definitions are, by the uniqueness of characters, equivalent.) This definition is unobvious, and worth closer examination. An individual,  $x$ , *has property E* iff there is a property  $F$  such that

- (i)  $F$  is a character of  $x$ , and,
- (ii)  $F$  is necessarily instantiated.

In symbols,

$$E[x] =_{df} \exists F(F[x] \ \& \ \forall H(H[x] \rightarrow \Box \forall y(F[y] \rightarrow H[y]) \ \& \ \Box \exists y F[y]).$$

(Gödel's definition is that any property  $F$  such that (i) is also such that (ii): in symbols

$$=_{df} \forall F((F[x] \ \& \ \forall H(H[x] \rightarrow \Box \forall y(F[y] \rightarrow H[y])) \rightarrow \Box \exists y F[y]).)$$

Note that, though both clause (i) and clause (ii) involve quantifying into the scope of necessity operators, they do so only with respect to Second-Order (property) variables.



The condition abbreviated as E[.] is therefore purely *de dicto*. I think it is just possible, in colloquial usage, to mean something like E in saying that something is a necessary being, or better that it is a necessary *sort of* being, but it is not the notion typically expressed by the words in academic philosophical usage. The sense is analogous to that in which you can point at the ants at a picnic and say ‘They are inevitable at picnics’—not that it is inevitable, *de re*, that the particular ants you are indicating show up at every picnic, but rather, *de dicto*, that visitation by creatures of the same sort as those indicated is inevitable.

Now we can state Gödel’s final axiom: necessary existence—E—is a positive property. I can’t say this seems a self-evident truth to me on a ‘moral aesthetic’ interpretation of positivity; as for pure ‘attribution’ as opposed to ‘privation,’ I note that there are no negation signs in the definiens for E, which may be some sort of evidence. Let us, however, go on with the formal argument, and consider later whether there is an interesting sense of positivity for which the axioms are all plausible.

Gödel’s final definition (in logical order—he actually states it first) is of the property of *godlikeness*, G. An individual is godlike iff it has every positive property,

$$G[x] =_{df} \forall F(P[F] \rightarrow F[x]).$$

Godlikeness, he notes, is a positive property: since we are not distinguishing between necessarily equivalent properties, G can be thought of as the conjunction of all positive properties, and so as positive by Gödel’s axiom about arbitrary conjunctions of positive properties. (Scott [23] does not include an axiom about infinite conjunctions, and simply takes the positivity of G as axiomatic, giving the argument that

[positivity] is a logical [feature of properties] and G is defined logically as an intersection of positive properties. Any such property ought also to be positive as an informal gloss on his axiom.)

Given that the positive properties form an ultrafilter on the Boolean algebra of properties, it is immediate that (it is necessarily true that) godlikeness is a character of any individual that has it. (This is why we didn’t have to worry about proving that all individuals have characters: it is enough that godlike ones do.) Since G is positive, it is possible that there is an individual with the character G:

$$\Diamond \exists x G[x].$$

Since E is a positive property, any possible individual with G will also have E: that is, its character will be necessarily instantiated:

$$\Box \forall (G[x] \rightarrow E[x]).$$

But G is the (only) character of such an individual, so it is possible that G is necessarily instantiated,

$$\Diamond \Box \exists x G[x].$$

by the principle (common to all normal modal logics) that necessary consequences of possibilities are possible. Whence S5 yields the actual, and indeed the necessary, existence of at least one godlike individual. The logic is unexceptionable, no axioms are required beyond those stated, and, as Adams [2] points out, the argument makes use only of *de dicto* modality (in the sense defended in the first section, in which 'quantifying in' with respect to property variables is not considered *de re*).

### III. Scott's Version

Gödel's [11] is short and cryptic. Gödel gave it to Dana Scott, and discussed the argument with him, early in 1970; Scott, before presenting the argument to a seminar in late 1970, prepared his own handout, [23]. It is, as those familiar with Scott's expositions of any number of topics in logic would expect, far clearer and easier to follow. Logically, Scott's argument follows Gödel's very closely, but Scott adds two additional bits. One is a gloss on the definition of property E (which he defines in Gödel's original, universally quantified, manner): Scott says that  $E[x]$  'means that  $x$  necessarily exists if it has [a character]'. Now, this makes no sense if the words 'necessarily exists' in it are taken in the defined sense of E: Scott must, therefore, be making a claim to the effect that, for individuals with characters (and he has stated no axioms implying that *all* individuals have characters), E implies necessary existence in some intuitive sense. The other addition, two lines earlier in [23] is (changing the notation to agree with ours)

$$F \text{ Ess. } x \rightarrow \Box \forall y (F[y] \rightarrow y = x).$$

Here 'F Ess.  $x$ ' means that F is a character (Scott follows Gödel in calling it an essence) of  $x$ , and the consequent is the explicitly *de re* claim that it is impossible for any individual other than  $x$  to have F. Does this follow from Gödel's axioms? It seems to me that it does not, if we adopt the more restricted form of property theory, with no constants or extraneous free individual variables allowed in the formula inserted in the comprehension scheme. On that version of the theory, Gödel's axioms seem compatible with a Leibnizian assumption that different possible worlds have disjoint domains, so that if a world other than the actual contains an individual with the character of Scott, it will be (not Scott, but) another individual of the same character. Even if we leave the necessity operator out of Scott's formula, so that it speaks only of actual individuals  $y$ , the formula expresses a strong 'identity of indiscernibles' thesis: each individual with a character is *uniquely* characterized by that character. Now, Leibniz, and Gödel as a Leibnizian, may have believed this, but it does not follow from the axioms stated, without assuming relational properties.

So perhaps Gödel, or at least Scott, was assuming the stronger form of property theory, with individual constants and extraneous variables allowed in comprehension axioms. On that theory, one of Scott's properties, and so one of the properties entailed by Scott's character (if he has a character) is that of *being identical to Scott*, and the formula follows from the axioms and definitions given. Further, when we assume haecceitistic properties, Scott's other comment has a natural interpretation: any individual with a necessarily instantiated character exists necessarily in the (now standard) *de re* sense:

$$x \text{ has a necessarily instantiated character} \rightarrow \Box \exists y (y = x).$$

Scott's additions to Gödel's argument, then, unlike Gödel's own conclusions, presuppose for their proof a notion of property whose acceptance is tantamount to accepting *de re* modality, and his second comment is, unlike anything in Gödel's own note, most naturally understood as an explicitly *de re* statement. (Adams [2] has already pointed out that Scott's argument, unlike Gödel's, does not avoid the *de re*.)

#### IV. Sobel's Criticism

I do not know how widely [11] and [23] circulated at the time.<sup>2</sup> According to Adams [2], Gödel told Oskar Morgenstern that, though he was 'satisfied' with the proof, he hesitated to publish it for fear it would be thought 'that he actually believes in God, whereas he is only engaged in a logical investigation . . .', so perhaps he discouraged circulation. (I find Morgenstern's testimony a bit mystifying, since Gödel *did* actually believe in God. Perhaps Morgenstern misunderstood him, and what he actually said was that he was afraid people would think he believed the proof *established* the existence of God, whereas he was only 'satisfied' that the argument *from certain axiomatized assumptions* to that conclusion could be shown to meet modern standards of rigor.) Neither note appeared in print until 1987, when Sobel included them as appendices to [25]. In the course of a very clear exposition and examination of the argument, Sobel noted that Gödel's axioms implied a further conclusion: all truths are necessary! Now, this might have been acceptable to Leibniz, who was a determinist, and perhaps to Gödel, who admired Leibniz and held the rationalistic view that whatever happens, happens for a reason, but Sobel and many others have found it objectionable.

The argument is quite simple. We have proven that there is, and necessarily must be, a godlike individual. Now consider an arbitrary true proposition, *P*. One of the properties the godlike individual (and every other individual) possesses is that of *being such that P is true*. Since the godlike individual has *only* positive properties, this property must be positive, and so entailed by godlikeness. So any world at which there is a godlike individual—any world at all, by Gödel's argument—will be a world at which at least one individual has this property. But if any individual has the property of being such that *P* is true, then *P* is true. So *P* is true at every world.

Given the version of modal property theory that countenances relational, haecceitistic, properties, this argument can be formalized for any proposition expressed by a sentence of our formal language, and requires no premises beyond those stated (in particular: no property-existence axioms beyond the comprehension principles). On the more restricted version of property theory, the conclusion only follows for *purely general* propositions: the restricted form of the comprehension principle does not allow us to prove there is such a property as being such that *P* is true if the proposition *P* is expressed by a sentence containing a proper name. On the restricted theory, then, it would not be possible to prove that it was a necessary truth that, say, Dana Scott presented a formulation of the ontological argument given him by an older logician to a seminar at one university and subsequently taught at another. It is, however, still possible to argue in Sobel's manner to

<sup>2</sup> In the late 1970s I had heard rumors that Gödel had shown a draft on the ontological argument to a colleague (and made a fool of myself by recounting the rumor to. . . Dana Scott!), but had no idea that [11] and [23] were in any sort of circulation.

establish the necessity of the existentially quantified proposition that *someone* presented a formulation of the ontological argument given him by an older logician to a seminar at one university and subsequently taught at another. And perhaps this is objectionable enough: we can prove that it is necessary that the world *look* the way it actually looks; the only contingent truths are those saying *which* individuals have which properties.

Probably Sobel's argument could be circumvented in many ways. Since the 'trick' in the argument turns on the fact that standard systems of Second-Order logic allow a proposition to be 'encoded' as a property of an individual even when the proposition is not, in any intuitive sense, *about* that individual, one promising avenue would be to investigate the consequences of a more restrictive theory of properties countenancing only properties in some sense *intrinsic* to their bearers. Such a theory might be developed along the lines of Dunn [7] and [8], but this would involve a radical departure from the purely modal logical framework of Gödel's (and, arguably, Leibniz's) reasoning.

#### V. Anderson's Emendation

Anderson [4] proposes what at first seems a more conservative modification of Gödel's argument to avoid Sobel's deterministic conclusion. Recall Gödel's 'ultrafilter' axiom: of any pair of a property and its negation, exactly one is positive. One point at which this was used in the proof was in showing that positive properties are possibly instantiated, but this argument required only that *at most* one member of the pair can be positive. Suppose we drop the other half of the axiom, the assumption that *at least* one member of any such pair is a positive property. Then the positive properties will still form a filter on the Boolean algebra of properties, but not necessarily an ultrafilter, and, similarly, the propositions 'encodable' by positive properties (even on the strong version of property theory countenancing relational properties) will form a filter, but not necessarily an ultrafilter, on the Boolean algebra of propositions. More intuitively, and recalling that 'theory' and 'complete theory' are the logical analogues of the algebraic notions 'filter' and 'ultrafilter', suppose that a godlike individual is still one with all positive properties, and that we can still prove that it is necessary that there be a godlike individual. This would imply that a certain theory (= set of propositions closed under entailment), namely, the theory consisting of propositions encoded by positive properties, is necessarily true. Since, however, there is no longer any reason to suppose that, from any pair of a proposition and its negation, at least one will be encoded by a positive property, this leaves an opening for significant amounts of contingency. It is worth investigating, therefore, whether the other half of the 'ultrafilter' axiom is required for the proof.

In fact, the assumption that at least one of any pair of a property and its negation is positive *was* used at one point: in proving that property G—godlikeness—would be a character of any individual that had it. (This, recall, was how we avoided the need for something like Sobel's axiom that individuals have characters.) But we don't want characters anyway: if we are to avoid Sobel's deterministic conclusion, we don't want to prove that any character is necessarily instantiated. We need to recast Gödel's argument, replacing the notion of a character with the notion of some kind of property logically weaker than a character.

Anderson proposes to use the notion of the *essence* of an individual, interpreting this notion in a thoroughly *de re*, fashion. An *essence* of an individual is, as we defined it

earlier, a property entailing all and only the properties that the individual has essentially; identifying mutually entailing properties this amounts to the conjunction of all the individual's essential properties. (Anderson uses a slightly different definition of essence; this is connected with his use of a 'free' logic of predication, and leads to complications we can ignore.) In the strong version of modal property theory, the comprehension principle suffices to prove that every individual has an essence: indeed, the haecceity of an individual is its essence! (Proof: Dana Scott's haecceity, to pick an arbitrary example, is the property of *being identical to Dana Scott*, and let us suppose that *rationality* is one of his essential properties. Suppose that some individual in some possible world possesses Scott's haecceity. Then this individual must be Scott, and so must have Scott's essential properties, such as rationality. So a haecceity entails all of its bearer's essential properties. On the other side, *Pennsylvania residence* is an accidental property of Scott's, so there is a world at which he lives in some place other than Pennsylvania, and at that world Scott's haecceity is co-instantiated with the negation of Pennsylvania residence, and, for properties like *stupidity* which he does not possess, the actual world is a world at which his haecceity is co-instantiated with their negations. Thus his haecceity does not entail any properties other than those he has essentially.)

An aside. Anderson seems happy with this strong property theory, and after this paragraph we will assume it for the remainder of this section, but the very triviality of the proof of the existence of essences it provides might be thought a mark against it. The essence of a thing, one might feel, ought to be an *informative* property of it. Now, our original reason for considering the restricted property theory was that allowing relational properties committed us, in effect, to allowing *de re* modal arguments, but the converse doesn't seem to hold: one could accept and argue in terms of *de re* modal notions but still accept only the restricted comprehension principles for properties—or, at the very least, still take an interest in a notion of non-relational property for which only the restricted and not the general comprehension principle is valid. The proposition that individuals have essences does not seem to be derivable from the comprehension principles in this theory, but would seem to be a plausible additional axiom, and given Anderson's definition of godlikeness it would be possible to prove, with no additional axioms, that godlikeness is an essence of any individual that has it. It would not be possible to prove in general that distinct individuals can't share an essence. Such a theory would seem in accord with Descartes's conception: his essence is simply to be a thinking thing, and (as you will discover if you think through the Second Meditation for yourself), so is yours. Another possibility would be to recognize both concepts of *property*, using two distinct alphabets of Second-Order variables, postulating the general form of the comprehension axioms for one and only the restricted form for the other. Defining an individual's essence as the conjunction of all its *non-relational* essential properties, it would be possible to prove on this basis that an individual's haecceity entails its essence (if it has an essence), but not the other way around. End of aside.

To base a proof on the notion of essence instead of that of character, Anderson gives new definitions of godlikeness,

$G^*[x] =_{df} x$  has *essentially* all (and only) the positive properties,

and of necessary existence,

$E^*[x] =_{df} x$  has a necessarily instantiated essence,

and postulates that  $E^*$ , rather than Gödel's  $E$ , expresses a positive property. With these changes, and a slight strengthening of the axioms (see comment toward end of section VI), the rest of the proof goes through as before.

Anderson notes that it is possible to prove, on his basis, not only that it is necessary that there be a godlike individual, but that this individual is unique. This, however, is simply Scott's addition to Gödel's argument, proven by appeal to the notion of a haecceity.

## VI. A *De Dicto* Variant

The key move in Anderson's emendation is the rejection of the 'ultrafilter' axiom. Without it, Gödel's notion of the character of an individual cannot be used in giving a proof, so the rest of Anderson's emendation consists in finding a substitute notion on which the conjunction of all positive properties can be shown to be the 'character-substitute' of a godlike individual. He chose the notion of the essence of an individual, and produced an argument which, unlike Gödel's, makes essential (if you will pardon the expression) use of *de re* modal notions. Another substitute is available, however, and allows a proof which, in its avoidance of modality *de re*, is perhaps closer in spirit to Gödel's own.

Define a property of an individual to be a *profile* of the individual if it entails all the positive properties that individual possesses, and entails no other properties:

$F$  is a profile of  $x =_{df} \forall H(\Box \forall x(F[x] \rightarrow H[x]) \leftrightarrow (P[H] \& H[x]))$ .

Assuming Gödel's principle that the positive properties are closed under arbitrary conjunctions, an individual's profile, therefore, is the conjunction of all and only that individual's positive properties: it is to positivity what an essence is to essentiality, or a character to properties in general. In what is becoming a familiar pattern, it can be proven that profiles are unique, but not, in general and on the basis of comprehension axiom alone, that every individual has a profile. A Sobel-style axiom asserting that every individual has a profile seems plausible enough, but won't actually be needed.

We may adopt Gödel's definition of godlikeness as the possession of all positive properties; it can be proven that godlikeness is the profile of any godlike individual. For necessary existence, we adopt a definition in terms of profiles:

$E^{**}[x] =_{df} \exists F(F \text{ is a profile of } x \& \Box \exists x F[x])$ ,

$x$  is a necessary existent ('a necessary sort of thing') if it has a necessarily instantiated profile. And we change the last axiom again, this time to say that  $E^{**}$  expresses a positive property. We may then argue as follows: as in Gödel's original proof, godlikeness is positive, and so possible. Thus there might be a godlike individual. Its (unique) profile would be godlikeness, and one of its properties would be that of having a necessarily instantiated profile. So it is possibly necessary, and so by S5 necessarily and actually the case, that there is a godlike individual.

There is one way in which our purely *de dicto* proof seems weaker than Scott's or Anderson's: we cannot prove that the godlike individual is unique on the basis of the axioms stated. A proof of the uniqueness of the godlike individual would depend on some further information about which properties were positive. The formal axioms allow, for example, an interpretation on which the universal property, self-identity, is the *only* positive property. (The last axiom, about the positivity of  $E^{**}$ , is the only even slightly problematic one: it amounts to the claim that no possible world has an empty domain.) On this interpretation, the conclusion reduces to the claim that  $\exists x(x = x)$ , asserting the satisfaction of a condition that is almost certainly satisfied by more than one individual! If, however, there is an interpretation on which godlike individuals would have properties such as are traditionally attributed to the Christian deity, it seems likely that an argument for uniqueness will be possible. Suppose, for example, that 'being the creator of every being distinct from itself' is a positive property. It is not possible for two individuals to have this property, so there could be at most one godlike individual. Similar remarks can be made about the *de re* necessary existence of a godlike individual. Nothing in the formal statement of the argument implies that the *same* godlike individual exists in every world: we have, after all, given a formulation designed specifically to *avoid* commitment to any non-trivial *de re* claims. If, however, godlikeness is interpreted in a way to make godlike individuals God-like, it becomes plausible to assume (and the 'counterpart theory' of Lewis [18], or any account of *de re* modality of even roughly similar spirit, would virtually force us to conclude) that if there is a godlike individual in every world, it is the same one.

There remains the question of plausibility. As suggested in the previous paragraph, the notion of positivity can be interpreted in ways that make the axioms plausible, but the conclusion trivial. It seems to me not impossible that there will be interpretation in which the proof would establish something interesting, but non-theological. For the original purpose of establishing the existence of God, however, I see my way clear to making only a relative claim: our version of the proof is no *worse* off than Gödel's original or Anderson's revision, and may in fact have weaker assumptions. Of the axioms, those according to which positivity is a non-contingent feature of properties, properties entailed by positive properties are positive, and arbitrary conjunctions of positive properties are positive, are common to all three proofs. All three assume further that the negation of a positive property is not positive; Anderson's and ours improve on Gödel's by dropping the 'ultrafilter' axiom.

Anderson, like Scott, takes as axiomatic the statement that godlikeness is a positive property, but, given the difference in definition, Anderson's axiom is stronger, and cannot be argued for simply on the basis that conjunctions of positive properties are positive. Anderson's  $G^*$  involves, not merely having all positive properties, but having them essentially; it is thus a stronger condition than Gödel's, and our,  $G$ , and the assertion that it expresses a positive property correspondingly stronger.

This leaves only the last axiom, that 'necessary existence,' defined as  $E$  by Gödel,  $E^*$  by Anderson, and  $E^{**}$  here, is a positive property. The essential properties of an individual and its positive properties are all properties of the individual: an individual's character (the character, that is, which it possesses at a world), therefore, is logically stronger than (i.e. entails) its essence or its profile (the profile it has at the same world). Thus if an individual's character (at a world) is necessarily instantiated, so are its essence

and its profile: in other words, necessary existence in Gödel's sense is a logically stronger condition than necessary existence in Anderson's or ours. But properties entailed by positive properties are positive: Gödel's axiom of the positivity of necessary existence entails (in the context of the other axioms) both Anderson's and ours, but not *vice versa*. So far, then, we have shown that Anderson's version and ours are *weaker* axioms than Gödel's.

Comparing Anderson's axiom to ours is more delicate, since there is no general agreement either about which properties are positive or about what the essential properties of individuals are. On plausible assumptions, however, it can be argued that positivity and essentiality (to most individuals) are independent. Certainly Kripke's thesis [16] about the essentiality of origins and original matter suggests that most of us have essential properties, such as being born to certain parents, or having started our existences composed of a certain quantity of matter, which do not seem likely to be positive on any theologically interesting interpretation of the latter notion: not all of our essential properties are positive. On the other hand, suppose that there is at least one positive property—*infinite wisdom*, perhaps—that not every individual possesses, and another, non-positive, property—perhaps *professorial rank*—possessed accidentally by some individual lacking the first. Then *being either infinitely wise or a professor* is a positive property (since it is entailed by the positive property wisdom) possessed accidentally by that individual: not all of that individual's positive properties are essential to it. These two counterexamples, however, are not fully parallel. The first involves logically very strong essential properties: it makes it seem likely that the essence of an individual is not in general a positive property. The second, however, involves only a weak, artificial, disjunctive positive property: by itself, it does not provide much evidence that the *profile* of an individual (its strongest positive property) is not essential to it, and there seem to be many interpretations of the notion of positivity on which it would be. The plausible but theologically useless interpretation suggested in our discussion of uniqueness, above, is one: self-identity is essential to every individual. At the other end of the scale, Christian metaphysicians willing to make *de re* modal claims have typically held that such salient properties of the deity as omnipotence, omniscience, benevolence, and so on—properties which might be thought positive in 'the moral aesthetic sense'—are essential to the deity. Formally, Anderson's definition, according to which a godlike individual has *all* positive properties essentially, would imply that, at least in the central case of a godlike individual, the individual's profile is one of its essential properties. On any of these grounds, then, the profile of at least a godlike individual would seem to be entailed by its essence, so the condition of having a necessarily instantiated profile (or: having, if godlike, a necessarily instantiated profile) would be weaker than that of having a necessarily instantiated essence. So Anderson's axiom of the positivity of  $E^*$  may be a stronger assumption than our axiom of the positivity of  $E^{**}$ .

## VII. A Note on Positivity

I will not attempt to give a substantial characterization of a notion of positivity: I have, myself, grave doubts that there is any such notion on which the axioms of (any of these versions of) the ontological proof are plausible and the conclusion theologically



interesting, but no convincing argument that there *couldn't* be such an interpretation. I will, however, consider what formal closure conditions might be postulated for the class of positive properties; my conclusions will suggest that it would be difficult to argue for the plausibility of some of the axioms in what are at least initially attractive ways.

Gödel suggests, as one possible way of understanding the notion of a positive property, the idea that it is one which involves pure 'attribution,' with no tinge of 'privation'. In a related note-book entry (printed in [12]), Gödel considers some of the formal consequences of attempting to implement this definition. Assume that the primitive predicates of a language express positive properties: then properties definable from them in certain ways ought also to be positive. Since Gödel considers only conditions on a single variable built up from monadic primitive predicates, he looks at forms of propositional composition. He suggests that a defined property is positive if, when its definiens is put into conjunctive normal form (that is, into the form of a conjunction of disjunctions of atomic and negated atomic formulas), each conjunct has at least one positive (i.e. un-negated) disjunct. He asserts the, not immediately obvious but readily verified, *theorem* that this is equivalent to counting as positive just those properties definable from the primitives through use of the connectives of disjunction, conjunction, and (material) implication: a pleasing syntactic criterion of positivity.

Now, even if one's preferred notion of positivity is a 'moral aesthetic' one, one might hope that the class of positive properties is closed under various sorts of definition, as Gödel asserts it is closed under arbitrary conjunction, and use the closure conditions to prove the positivity of certain defined properties. Unfortunately, Gödel's little theorem about the propositional connectives that can be used in defining positive properties from positive properties is not enough. The key defined properties figuring in the various versions of Gödel's ontological argument, G, Anderson's G\*, E, E\*, and E\*\*, all have definitions that involve quantifiers and modal operators as well as truth-functional connectives.

First the quantifiers. (In order to make use of First-Order quantifiers in defining new positive properties, we must either allow relational properties or extend the notion of positivity to apply to relations as well as monadic properties, and enquire into the conditions under which quantifying into an argument place of a positive  $n+1$  place relation yields a positive  $n$  place propositional function.) Existential quantification is clearly legitimate: existentially quantifying yields a weaker condition, and any property entailed by a positive property is positive. Universal quantification is more problematic: as Adams remarks, 'as any logician knows, universal quantifiers are equivalent to a sort of negation' [1, p. 116]. To warm up with a small problem about First-Order universal quantification, suppose that identity is a positive notion (we will consider below whether this is plausible). Then, if properties defined from positive relations by universal quantification are positive, the property

$$U[x] =_{df} \forall y (y = x)$$

will be positive. But only an extreme monist (Spinoza?) would want the ontological argument to yield the existence of an individual with *this* property! (The example is suggested by a criticism Leibniz made of an attempt by Eckhard to define the notion of an *ens perfectissimum*—cf. [1, p. 131].) Second-Order universal quantification is even more

problematic, since, in combination with material implication, it yields a well-known definition of negation:

$$\neg F[x] =_{\text{df}} \forall H(F[x] \rightarrow H[x]).$$

This is very disturbing, since godlikeness is defined as the possession of *all* positive properties. Perhaps certain, restricted, uses of Second-Order universal quantification are legitimate. It is possible to have doubts about the use of the notion of positivity itself in the definition of positive properties—positive properties, after all, are characterized in part as those which do *not* imply any imperfection or privation—but it seems plausible to allow it and Second-Order universal quantification in combination: a property defined from a positive one by universal quantification *restricted* to positive properties (a property, that is, defined by a condition of the form  $\forall F (P[F] \rightarrow \dots)$ , where the condition replacing the dots is positive) seems likely to be positive. Such, at least, seems to have been Scott's intuition, given his comment on the definition of godlikeness, which is of this form. (If you are really worried by the appearance of the positivity predicate in the definiens, note that a condition defined by such restricted universal quantification of an otherwise positive condition is equivalent—in the modal sense of mutual entailment—to an infinitary conjunction—one conjunct for each positive property—of positive conditions.)

As an aside, this suggests that *identity of individuals* may not be a positive notion, if it is defined in the Leibnizian fashion

$$x = y =_{\text{df}} F\forall(F[x] \leftrightarrow F[y]),$$

for (unless we adopt Gödel's ultrafilter assumption) the universal quantifier in this definition cannot be restricted to positive properties.

As for the modal operators, the possibility operator is unproblematic, for the same reason as given for existential quantification. I have no firm intuition about the necessity operator, but its analogies with (Second-Order) universal quantification make me cautious. Unfortunately, it is precisely the use of the necessity operator in defining positive properties that we would need to get clearer about if we were to have any hope of proving the positivity of necessary existence (on any of the three proposed definitions of this notion).

University of Melbourne

Received: November 1997.

Revised: February 1998.

# CORRIGENDUM

The schema proposed on page 366 is reasonable for “filter-like” conditions such as the characterhood, essencehood, and profilehood considered in the paper. It fails, however, in the case of a condition (such as, e.g., “holds of fewer objects than not”) false of some property entailed by a conjunction of properties it is true of. A generally valid infinitary conjunction schema is obtained by postulating, for each condition, a property which (i)

entails every property satisfying the condition, and (ii) is itself entailed by any other property entailing every property satisfying the condition.

## REFERENCES

1. Robert M. Adams, *Leibniz: determinist, theist, idealist* (New York: Oxford University Press, 1994).
2. —, Introduction to Gödel [11] in Gödel [12].
3. Alan Ross Anderson and Nuel D. Belnap, *Entailment*, Vol. I, (Princeton: Princeton University Press, 1975).
4. C. Anthony Anderson, 'Some emendations of Gödel's ontological proof', *Faith and Philosophy* 7 (1990), pp. 291–303.
5. Aldo Bressan, *A General Interpreted Modal Calculus* (New Haven: Yale University Press, 1972).
6. Rudolf Carnap, *Logical Syntax of Language* (London: Routledge & Kegan Paul, 1937); expanded Eng. tr. of work published in German in 1934.
7. J. Michael Dunn, 'Relevant predication I: the formal theory', *Journal of Philosophical Logic* 16 (1987), pp. 347–381.
8. —, 'Relevant predication II: intrinsic properties and internal relations', *Philosophical Studies* 60 (1990), pp. 177–206.
9. Kit Fine, 'Model theory for modal logic, part II: the elimination of *de re* modality', *Journal of Philosophical Logic* 7 (1978), pp. 277–306.
10. Daniel Gallin, *Intensional and Higher-Order Modal Logic* (Amsterdam: North-Holland, 1975).
11. Kurt Gödel, 1970 note on the ontological argument given to Dana Scott; published as an appendix to [25] and, along with related manuscript material, in [12].
12. —, *Collected Works*, Vol. III, edited by Solomon Feferman *et al.* (New York: Oxford University Press, 1995).
13. A.P. Hazen, 'The elimination of *de re* modalities', unpublished typescript circulated early 1976.
14. —, 'Actualism again', *Philosophical Studies* 84 (1996), pp. 155–181.
15. Saul Kripke, 'Identity and necessity', in *Identity and Individuation*, ed. M. Munitz, (New York: New York University Press, 1971); repr. in *The Possible and the Actual*, ed. M.J. Loux, (Ithaca: Cornell University Press, 1979).
16. —, 'Naming and necessity', in *Semantics of Natural Language*, eds. D. Davidson and G. Harman (Dordrecht: Reidel, 1972); repr., with a new introduction, as Saul Kripke, *Naming and Necessity* (Cambridge: Harvard University Press, 1980).
17. —, review of Fine [9] (and of parts I and III), *Journal of Symbolic Logic* 50 (1985), pp. 1083–1093.
18. David K. Lewis, 'Counterpart theory and quantified modal logic', *Journal of Philosophy* 65 (1968), pp. 113–126; repr. in *The Possible and the Actual*, ed. M.J. Loux (Ithaca: Cornell University Press, 1979), and with additions in D.K. Lewis, *Philosophical Papers*, Vol. I (New York: Oxford University Press, 1983).
19. —, 'New work for a theory of universals', *Australasian Journal of Philosophy* 69 (1983), pp. 343–377.
20. —, *On the Plurality of Worlds* (Oxford: Basil Blackwell, 1986).
21. Benson Mates, *The Philosophy of Leibniz* (New York: Oxford University Press, 1986).
22. Terence Parsons, 'Essentialism and quantified modal logic', *Philosophical Review* 78 (1969), pp. 35–52.
23. Dana Scott, 1970 notes on Gödel's argument, published as an appendix to Sobel [25].
24. Alan Sidelle, *Necessity, Essence and Individuation: a defense of conventionalism* (Ithaca: Cornell University Press, 1989).
25. Jordan Howard Sobel, 'Gödel's ontological proof', in *On Being and Saying: essays for Richard Cartwright*, ed. Judith Jarvis Thompson, (Cambridge: MIT Press, 1987).
26. Peter F. Strawson, *Individuals* (London: Methuen, 1959).
27. Richmond H. Thomason, 'Modal logic and metaphysics', in *The Logical Way of Doing Things*, ed. Karel Lambert, (New Haven: Yale University Press, 1969).