10

A Note on Kripke's Puzzle about Belief

Nathan Salmon

Millianism is the doctrine that the semantic content of a proper name is just the name's designatum. Without endorsing Millianism Kripke uses his well-known puzzle about belief as a defense of Millianism against the standard objection from apparent failure of substitution. On the other hand, he is not resolutely neutral. Millianism has it that Pierre has the contradictory beliefs that London is pretty and that London is not pretty – that Pierre both believes and disbelieves that London is pretty. I argue here for hard results in connection with Saul Kripke's puzzle and for resulting constraints on a correct solution. Kripke flatly rejects as incorrect the most straightforwardly Millian answer to the puzzle. Instead he favors a view according to which not all instances of his disquotational principle schema and its converse (which taken together are equivalent to his strengthened disquotational schema) are true although none are false. I argue in sharp contrast that the disquotational schema is virtually analytic. More accurately, every instance of the disquotational schema (appropriately restricted) is analytic. Moreover, there is an object-theoretic general principle that underlies the disquotational schema, is itself analytic, and entails each of the instances of the disquotational schema. By contrast, the converse of the disquotational principle leads to a genuine contradiction and is thereby straightforwardly falsified by Kripke's own example.

Much of the present material was presented in a nutshell to the Gala Opening of the Saul Kripke Center at the Graduate Center of the City University of New York on May 21, 2008. I am grateful to my audience, especially Saul Kripke, for their reactions and comments. I also thank David Kaplan for prior discussion. I am equally grateful to the participants in my seminar at the University of Southern California during fall 2008, especially Daniel Kwon and Lewis Powell, for their many helpful comments.

I

I argue here for relatively hard results in connection with Saul Kripke's well-known puzzle about belief, and for resulting constraints on a correct solution.[1] Kripke uses the puzzle as part of a defence of Millianism against the standard objection from apparent failure of substitution. He does not endorse Millianism, however. Indeed, he is not even resolutely neutral, for he also flatly rejects as incorrect the most straightforwardly Millian answer to the puzzle. The defense consists in exposing that the traditional objection from substitution failure implicitly invokes a set of supplementary assumptions that, by themselves, generate the same counterintuitive consequence completely independently of Millianism (pp. 1018–19).

In presenting the puzzle, Kripke follows a sound methodology championed in Alfred Tarski's classic discussion of the liar paradox ("antinomy"). Tarski wrote:

> In my judgment, it would be quite wrong and dangerous from the standpoint of scientific progress to depreciate the importance of this [the liar paradox] and other antinomies, and to treat them as jokes or sophistries. It is a fact that we are here in the presence of an absurdity, that we have been compelled to assert a false sentence.... If we take our work seriously, we cannot be reconciled with this fact. We must discover its cause, that is to say, we must analyze premises upon which the antinomy is based; we must then reject at least one of these premises, and we must investigate the consequences which this has for the whole domain of our research.[2]

In this same scientific spirit, Kripke enumerates each of the assumptions involved in obtaining the unacceptable conclusion, in order to identify and isolate the faulty assumption. There is to begin with the *principle of translation*:

> *T*: If a sentence of one language expresses a truth in that language, then any literal (that is, semantic-content-preserving) translation of it into any other language also expresses a truth in that other language.[3]

[1] "A Puzzle about Belief," in A. Margalit, ed., *Meaning and Use* (Dordrecht: D. Reidel, 1979), pp. 239–83; reprinted in M. Davidson, ed., *On Sense and Direct Reference* (Boston: McGraw Hill, 2007), pp. 1002–36. Page references throughout are to this reprinting.

[2] Tarski, "The Semantic Conception of Truth," *Philosophy and Phenomenological Research*, 4 (1944), pp. 341–75, at p. 348; reprinted in L. Linsky, ed., *Semantics and the Philosophy of Language* (Urbana: University of Illinois Press, 1952), pp. 13–47, at p. 20.

[3] I have inserted the phrase 'literal (that is, semantic-content-preserving)' on Kripke's behalf. This notion of literal translation is clearly intended, both by Alonzo Church and by Kripke; nonliteral translation is irrelevant. See my "The Very Possibility of Language," in C. A. Anderson and M. Zeleny, eds., *Logic, Meaning and Computation: Essays in Memory*

At the center of the puzzle is the *disquotational schema* for English:

> $D_{English}$: If a normal English speaker, on reflection and under normal circumstances, sincerely assents to 'φ' then he/she believes that φ.[4]

Infinitely many disquotational principles are thus obtained by replacing both occurrences of the schematic letter 'φ' "by any appropriate standard English sentence lacking indexical or pronominal devices or ambiguities" (p. 1014).

There is an analogous French schema D_{French} for French, an analogous Italian schema $D_{Italian}$ for Italian, and so on. The translation of D_{French} into English is the following schema, where 'φ_F' is to be replaced (within the quotation marks) by any appropriate standard French sentence lacking indexical or pronominal devices and 'φ_E' by that sentence's translation into English:

> D_{French}*: If a normal French speaker, on reflection and under normal circumstances, sincerely assents to 'φ_F' then he/she believes that φ_E.

There is also a strengthened disquotational schema for English:

> $SD_{English}$: A normal English speaker who is not reticent will be disposed under normal circumstances to sincere reflective assent to 'φ' if and only if he/she believes that φ. (p. 1015)

of Alonzo Church (Boston: Kluwer, 2001), pp. 573–95; reprinted in my *Metaphysics, Mathematics, and Meaning* (Oxford: Oxford University Press, 2005), chapter 17, pp. 344–64, at 352–4.

It must be noted that *T* extends to attributions of belief. Thus, if 'Pierre croit que Londres est jolie' is true in French then its literal translation into English is equally true in English. Assuming that the normal translation preserves semantic content, if 'Pierre croit que Londres est jolie' is true in French then 'Pierre believes that London is pretty' is true in English. Kripke demonstrates with his Paderewski example that this assumption is not essential to the puzzle.

[4] Kripke's formulation omits the phrase "under normal circumstances." I regard this as a minor oversight. Kripke says the following of his formulation: "I fear that even with all this some astute reader – such, after all, is the way of philosophy – may discover a qualification I have overlooked, without which the asserted principle is subject to counterexample. I doubt, however, that any such modification will affect any of the uses of the principle to be considered below. Taken in its obvious intent, after all, the principle appears to be a self-evident truth" (pp. 1014–15). The phrase "normal circumstances" is to be understood so that Pierre's inability to translate 'London' as 'Londres' or vice versa does not in itself disqualify his circumstances from being normal. Rather, the spirit of the principle schema excludes genuinely bizarre circumstances – as, for example, in which a normal speaker is under a hypnotic spell, or under the control of a Cartesian demon, and as a result signals assent when he/she intends to dissent.

(See note 4.) This strengthening of $D_{English}$ leads to a stronger form of the puzzle.

Of the various principles just enumerated, the translation principle *T* is the most immune from reasonable doubt. It is surely not the culprit. Indeed, Kripke demonstrates with his Paderewski example that *T* plays no crucial role in the puzzle. (See note 3.) I shall simply assume *T* throughout the present discussion.

To construct the puzzle, Kripke describes a hypothetical scenario in which a bilingual English-French speaker, Pierre, is unaware that the cities he calls 'London' and 'Londres' are in fact one and the same. The puzzle is generated through application of these various principles to the following stipulations taken to be true *by hypothesis*:

H1: Pierre is a normal French speaker.
H2: Pierre is a normal English speaker.
H3: Pierre is rational/logical.
H4: When confronted with 'Londres est jolie', 'London is pretty', or their negations, Pierre reflectively interprets the sentence.
H5: Pierre is not reticent to reveal his position with respect to the issue of whether London is pretty.
H6: Under normal circumstances Pierre sincerely assents to 'Londres est jolie' and is not at all disposed to assent to 'Londres n'est pas jolie'.
H7: Under normal circumstances Pierre sincerely assents to 'London is not pretty' and is not at all disposed to assent to 'London is pretty'.

We may supplement these stipulative hypotheses with the following trivially true hypotheses:

H8: English is a language; French is a language; 'London is pretty' and 'London is not pretty' are commonplace English sentences, which express as their English semantic contents, respectively, that London is pretty and that London is not pretty; and 'Londres est jolie' is a commonplace French sentence, which expresses as its French semantic content that London is pretty.[5]
H9: 'Pierre believes that London is pretty', 'Pierre does not believe that London is pretty', and 'Pierre believes that London is not pretty' are commonplace English sentences, which express as

[5] Normal French speakers (in Kripke's sense) inform me that '*jolie*' correctly applies in idiomatic French to a creature, not to a city. I shall follow the literature in ignoring this departure.

their English semantic contents, respectively, that Pierre believes that London is pretty, that Pierre does not believe that London is pretty, and that Pierre believes that London is not pretty; and 'Pierre croit que Londres est jolie' is a commonplace French sentence, which expresses as its French semantic content that Pierre believes that London is pretty.

Hypothesis *H8* has the straightforward consequence that the French sentence 'Londres est jolie' translates into English literally (preserving semantic content) as 'London is pretty', and vice versa. Hypothesis *H9* has the consequence that the French 'Pierre croit que Londres est jolie' translates into English literally as 'Pierre believes that London is pretty'. Again, Kripke demonstrates through his Paderewski example that these consequences concerning literal translation are in any event at least largely inessential to the puzzle.[6]

Kripke's puzzle presses a pair of questions:

Q1: Does Pierre believe that London is pretty?
Q2: Does Pierre disbelieve that London is pretty?

To *disbelieve* a proposition *p* is (at least for present purposes) to believe its negation, *~p*. It is thus a kind of believing.[7] The second question is thus whether Pierre believes that London is *not* pretty.

The preceding presentation is more explicit than Kripke's. Kripke focuses almost exclusively on *Q1*, though *Q2* is equally relevant and, strictly speaking, a distinct question from *Q1* (which might even be answered independently of *Q1*). More significantly, in presenting the puzzle Kripke avoids talk of propositions nearly altogether. This is not because he disbelieves in propositions or is skeptical of their existence. Rather he wishes to rest the puzzle on as meager resources as possible. Yet acknowledgment of propositions is at least implicit in the puzzle, and is crucial to solving it. Indeed, talk of propositions is already explicit at least in hypothesis *H4*, if not also in *H5*, neither of which does Kripke

[6] One purported solution that is Fregean in spirit (although it deviates significantly from Frege's own theory) has the implausible consequence that whereas 'Londres est jolie' translates literally into English as 'London is pretty', contrary to *H9* 'Pierre croit que Londres est jolie' does not translate literally as 'Pierre believes that London is pretty', and is translatable literally only insofar as the sense of 'Londres' in Pierre's French idiolect is expressible in English, as perhaps by a definite description. (See note 3.) Any purported solution is discredited to the extent that it is committed to rejecting trivial hypotheses.

[7] The interrelationships among belief, disbelief, suspension of judgment, and failure to believe are significantly more complicated than might first appear. See my "Being of Two Minds: Belief with Doubt," *Noûs*, 29, 1 (January 1995), pp. 1–20; reprinted in my *Content, Cognition, and Communication* (Oxford: Oxford University Press, 2007), pp. 230–48.

explicitly and fully state as such. If I am correct, as we shall soon see, talk of propositions is implicit also in *H1*, *H2*, *H6*, and *H7*. More important, reference to propositions, as we shall also see, underlies the disquotational schemata.

The relevant instances of $D_{French}{}^*$ and $D_{English}$ are the following:

$D_{French}{}'$: If Pierre is a normal French speaker and, on reflection and under normal circumstances, he sincerely assents to ‘Londres est jolie’, then he believes that London is pretty.

$D_{English}{}'$: If Pierre is a normal English speaker and, on reflection and under normal circumstances, he sincerely assents to ‘London is not pretty’, then he believes that London is not pretty.[8]

Invoking *H1*, *H2*, *H4*, *H6*, *H7*, *H8*, and *H9*, one obtains the bizarre result that Pierre both believes and disbelieves that London is pretty:

R1

a:	D_{French}, *T*	⊢	$D_{French}{}^*$.
b:	$D_{French}{}^*$, *H1*, *H4*, *H6*, *H8*, *H9*	⊢	Pierre believes that London is pretty.
c:	$D_{English}$, *H2*, *H4*, *H7*, *H8*	⊢	Pierre disbelieves that London is pretty.
d:	$D_{English}$, D_{French}, *T*, *H1*, *H2*, *H4*, *H6–H9*	⊢	Pierre has contradictory beliefs.

The weaker disquotational schemata together with *T*, the stipulative hypotheses, and the trivial hypotheses thus yield affirmative answers to both *Q1* and *Q2*. The primary version of the puzzle presses the obvious objection: Answering both questions affirmatively is evidently incompatible with *H3*. This conflict casts serious doubt on the disquotational schemata.

There is worse yet to come. The relevant instance of $SD_{English}$ is the following:

[8] Strictly speaking, these do not qualify as admissible instances, in light of the ambiguity in English of ‘London is pretty’, which can be used to describe London, Ontario, instead of London, England. The example illustrates that the disquotational schemata can be extended to ambiguous sentences, provided the sentence in question is given the same reading in the metalanguage that the speaker gives it in assenting or not assenting to it.

On the other hand, the disquotational schemata need to be restricted to sentences that are commonplace – that is, not technical, not especially long, with no arcane vocabulary, and so on. (See the preceding note.) For Kripke’s purposes, what are needed are plausibly restricted schemata for which $SD_{English}{}'$ and $SD_{French}{}'$ qualify as legitimate instances.

$SD_{English}'$: If Pierre is a normal English speaker and not reticent, then he will be disposed under normal circumstances to sincere reflective assent to 'London is pretty' iff he believes that London is pretty.

Invoking *H5* in combination with the same hypotheses as before, one now obtains results that are not merely implausible or mysterious, but utterly unacceptable – for example, that Pierre both believes, and also does not believe, that London is pretty.[9] In the spirit of Tarski's analysis of the liar paradox, Kripke's puzzle generates a fundamental result, which any solution must accommodate:

R2

a:	*H1, H2, H4–H9*	$\vdash$	$\neg(SD_{English} \wedge D_{French} \wedge T)$.[10]
b:	*H1, H2, H4–H9, T*	$\vdash$	$\neg(SD_{English} \wedge D_{French})$.

Assuming *T* together with the hypotheses listed, this result excludes the prospect that all instances of the strengthened disquotational schemata are true. On the other hand, it leaves open the issue of whether the weaker disquotational schemata might yet obtain.

II

Unlike Tarski, Kripke does not make any official pronouncement concerning which principles are guilty. Instead he considers a variety of possible answers to the puzzle without officially endorsing any of them. He does clearly favor one answer to the puzzle and flatly rejects some specific answers as incorrect – including the most straightforwardly Millian conclusion, to wit, that Pierre indeed has contradictory beliefs. Kripke objects that, given *H3*, "it is clear that Pierre, as long as he is unaware that the cities he calls 'London' and 'Londres' are one and the same, is in no position to see, by logic alone, that at least one of his beliefs must

[9] Kripke presents a third version of the puzzle on which *H7* is replaced with the following:

> ***H7'***: Pierre is not at all disposed to assent to either 'London is pretty' or 'London is not pretty'; instead his attitude is one of suspension of judgment. (p. 1022)

This replacement leads equally to the unacceptable conclusion that Pierre both believes and does not believe that London is pretty.

[10] The negation sign '¬', as contrasted with '~', will be used throughout to indicate that not all instances of the schema to which the sign is prefixed are true. (It does not in general follow that all instances of the schema are false, or even that any are.)

be false. He lacks information, not logical acumen. He cannot be convicted of inconsistency; to do so is incorrect" (p. 1022).

The solution Kripke favors accepts the translation principle *T*, but does not accept all admissible instances of any of the disquotational schemata. At the same time the favored solution does not reject any instance as false. Instead all problematic instances – Pierre vis-à-vis 'London is not pretty' and 'Londres est jolie', the ancients vis-à-vis 'Hesperus appears in the evening sky' and 'Phosphorus does not appear in the evening sky', Lois Lane vis-à-vis 'Superman can fly' and 'Clark Kent cannot fly', and so on – are deemed not true but also not false. On this solution it is neither true nor false that the ancients believed that Hesperus was Phosphorus, and neither true nor false that they believed that Hesperus was not Phosphorus. Analogously, on this solution it is neither true nor false that Lois Lane believes that Superman can fly, and neither true nor false that she believes that he cannot fly. And likewise, it is allegedly neither true nor false that Pierre believes that London is pretty, and neither true nor false that he believes that London is not pretty. Instead the phrase 'believes that', and perhaps even the simple proposition-designation forming operator 'that' by itself, are evidently undefined for these notorious problem cases. In the preface to *Naming and Necessity* Kripke writes:

> Some critics of my doctrines, and some sympathizers, seem to have read them as asserting, or at least implying, a doctrine of the universal substitutivity of [codesignative] proper names. This can be taken as saying that a sentence with 'Cicero' in it expresses the same 'proposition' as the corresponding [result of substituting 'Cicero'] with 'Tully', that to believe the proposition expressed by the one is to believe the proposition expressed by the other, or that they are equivalent for all semantic purposes. Russell does seem to have held such a view for 'logically proper names', and it seems congenial to a purely 'Millian' picture of naming, where only the referent [designatum] of the name contributes to what is expressed. But I.... never intended to go so far. My view that the English sentence 'Hesperus is Phosphorus' could sometimes be used to raise an empirical issue while 'Hesperus is Hesperus' could not shows that I do not treat the *sentences* as completely interchangeable. Further, it indicates that the mode of fixing the reference is relevant to our epistemic attitude toward the sentences expressed.[11] How this relates to the question what 'propositions' are expressed by these sentences, whether

[11] Our epistemic attitude toward the *sentences expressed*? Is this a slip of the pen? Sentences are not expressed *by* anything; sentences express propositions. Does Kripke mean our cognitive attitude toward *what* the sentences express, that is, toward the *propositions* that the sentences express? Does he mean our cognitive toward the *sentences* themselves – as opposed to the propositions they express? Neither? Both?

> these 'propositions' are objects of knowledge and belief, and in general, how to treat names in epistemic [i.e., propositional-attitude] contexts, are vexing questions. I have no 'official doctrine' concerning them, and in fact I am unsure that the apparatus of 'propositions' does not break down in this area. [*Footnote*: Reasons why I find these questions so vexing are to be found in my 'A Puzzle about Belief'.] (pp. 20–1)

In "A Puzzle," Kripke voices his worries in a similar manner:

> The point is *not*, of course, that codesignative proper names *are* interchangeable in belief [propositional-attitude] contexts *salva veritate*, or that they *are* interchangeable in simple contexts even *salva significatione*. The point is that the absurdities that disquotation plus substitutivity would generate are exactly paralleled by absurdities generated by disquotation plus translation, or even 'disquotation alone' (or: disquotation plus homophonic translation)…. When we enter into the area exemplified by … Pierre, we enter into an area where our normal practices of interpretation and attribution of belief are subjected to the greatest possible strain, perhaps to the point of breakdown. So is the notion of the *content* of someone's assertion, the *proposition* it expresses. (pp. 1033–4)

It can be forcefully argued – and I am persuaded – both that the solution Kripke favors is incorrect and furthermore that the answer he rejects is in fact correct. Pierre does indeed have contradictory beliefs. Whereas believing contradictions is typically a violation of even the most lenient of reasonable cognitive norms, in Pierre's circumstances the transgression is completely excused. What are at issue are precisely the weaker disquotational instances D_{French}' and $D_{English}'$. Whereas Kripke is inclined to deem them neither true nor false, because not all instances are true even if none are false, it can be demonstrated that they may be plausibly interpreted in such a way that they are basically *analytic* – or at least nearly enough so that they are straightforwardly true – even while hypotheses *H1*–*H7*, understood correspondingly, remain true by stipulation.

To substantiate the case, I shall propose definitions for 'normal speaker', 'reflect', 'sincere assent', and 'reticent', as these terms arise in Kripke's puzzle. In proposing these definitions I am guided by Kripke's own clarifications (at p. 1014). The point of these proposed definitions is not to capture the terms' standard English meanings. The point, rather, is to provide a set of concepts – core *potential* meanings – that are not implausible as contents for the terms, and that play the roles of such concepts as that of *normal speaker*, *sincere assent*, and so on in a fruitful reformulation of Kripke's puzzle, with the result that relevant speculative principles and stipulated hypotheses, so interpreted, are more readily assessed as legitimate or not.

The definition for 'normal speaker' is straightforward:

D1: Agent *A* *speaks* language *L* *normally* $=_{def}$ *A* speaks *L* sufficiently well that for every commonplace expression of *L*, *A* would normally use and take it to mean exactly what the expression in fact means in *L*; in particular, for every commonplace sentence *S* of *L*, if confronted with *S*, *A* would normally take it to express exactly the very proposition p_S that *S* in fact expresses in *L*.[12]

The issues surrounding the notions of reflection and sincerity are more complex and require more careful consideration. We begin by considering the following natural definitions as at least first approximations:

D2: Agent *A* *reflects with respect to* sentence *S* $=_{def}$ *A* considers *S* sufficiently that he/she thereby interprets it as expressing exactly the proposition p_{AS} he/she would normally take it to express.

D3: Agent *A* *sincerely assents to* sentence *S* $=_{def}$ *A* assents verbally to *S*; furthermore *A*'s verbal assent to *S* is appropriately occasioned by *A*'s believing p_{AS}, where p_{AS} is the very proposition he/she therewith takes *S* to express.

These definitions are in fact better than mere approximations. They closely reflect Kripke's own explanations of the relevant notions.[13] They also suffice, when taken in conjunction with *D1* and the stipulative and trivial hypotheses, for the purpose of establishing, contrary to Kripke himself, that Pierre indeed harbors contradictory beliefs. We shall consider a variety of possible refinements of *D2* and *D3*, but for simplicity's sake we shall take these to be our official definitions.

A more subtle pair of concepts is available, and equally sufficient for the purpose at hand. It might be supposed that the relevant notion of sincere assent essentially involves taking a metaperspective, specifically, taking oneself to believe the proposition expressed by the sentence to which one assents. Correspondingly, the relevant notion of reflection would involve getting oneself right. As such the following alternative definitions might be taken in lieu of *D2* and *D3*:

[12] See note 8. Kripke says, "When we suppose that we are dealing with a normal speaker of English, we mean that he uses all words in the sentence in a standard way, combines them according to the appropriate syntax, etc.: in short, he uses the sentence to mean what a normal speaker should mean by it" (p. 1014).

[13] Kripke: "The qualification 'on reflection' guards against the possibility that a speaker may, through careless inattention to the meaning of his words or other momentary conceptual or linguistic confusion, assert something he does not really mean, or assent to a sentence in linguistic error. 'Sincerely' is meant to exclude mendacity, acting, irony, and the like" (p. 1014).

D2′: Agent *A reflects with respect to* sentence $S =_{def} A$ considers *S* sufficiently that he/she thereby interprets it as expressing exactly the proposition p_{AS} he/she would normally take it to express; furthermore, in so doing *A* considers p_{AS} sufficiently thoroughly that, under normal circumstances, if he/she takes him/herself to believe p_{AS}, to disbelieve p_{AS}, or to suspend judgment, he/she so takes him/herself appropriately precisely because he/she does so believe, disbelieve, or suspend judgment.

D3′: Agent *A sincerely assents to* sentence $S =_{def} A$'s assent to *S* is appropriately occasioned by *A*'s *taking him/herself to believe* p_{AS}, where p_{AS} is the very proposition he/she therewith takes *S* to express.

Compared to these alternative definitions, the notions of reflection and sincere assent captured in *D2* and *D3* are somewhat crude. Although they yield cruder notions, the original definitions seem entirely faithful to Kripke's expressed intent. More important, the notions captured in *D2′* and *D3′* complement each other in such a way that the net effect of the replacements leaves the puzzle and the constraints on its correct solution exactly the same as with the cruder notions they replace.

Reflection on each of these various definitions confirms that each of the stipulated hypotheses *H1*–*H7*, as thus interpreted, may be taken to be true by hypothesis. For example, it may be taken as stipulated that if confronted with any commonplace French sentence, Pierre would normally take the sentence to express the very proposition it in fact expresses in French. Similarly for *H2* and each of *H4*–*H7*.

One might hesitate over *H4*. The proposed definition *D2′* defines 'reflection' in such a way that if an agent "reflects" with respect to a sentence, and judges under normal circumstances that he/she believes the proposition he/she therewith interprets the sentence as expressing, this is precisely because he/she does believe the proposition. Is it really legitimate simply to stipulate that in the scenario under consideration, if Pierre takes himself to believe the very proposition he interprets a sentence as expressing then he takes himself correctly? Or does such a verdict simply beg the question?

Controversy about Pierre's case notwithstanding, the stipulation must be deemed entirely legitimate. First, *H4* does not by itself settle the issue raised by *Q1* or *Q2* any more than any other enumerated hypothesis does; *H4* is no more question-begging, in any significant sense, than any of the other hypotheses are. It is also important to recognize exactly what *H4* stipulates. Interpreted through *D2′*, *H4* does not amount to a claim that Pierre is infallible concerning whether he believes. It does not even stipulate that Pierre is immune from error, that he *could not* be mistaken, in judging that he has a certain opinion. It stipulates a truth-functional relation: either Pierre does not judge that he believes, disbelieves, or suspends judgment about something, or else he does so judge but not under normal circumstances, or else he does so judge under normal circumstances and in so doing he considers matters sufficiently thoroughly so that in those circumstances his judgment is not mistaken. Taking oneself to be of a certain opinion is not like taking oneself to be healthy, wealthy, and wise. Careful, thoughtful, and thorough consideration normally provides considerably greater warrant, and greater likelihood of being correct, in the former case than in the latter. In judging that one is indeed of a certain opinion, for one to base that judgment on a cold, hard look at oneself in a careful and probing way is for one to examine thoroughly all the relevant evidence available. Normally, if one thoroughly considers the question of whether one believes something, and concludes that one does indeed believe, that conclusion is not merely a coincidently correct conviction. It is normally a firm case of knowledge. *D2′* might even be revised as follows:

> D2″: Agent *A reflects with respect to* sentence $S =_{def} A$ considers S sufficiently that he/she thereby interprets it as expressing exactly the proposition p_{AS} he/she would normally take it to express; furthermore, in so doing *A* considers p_{AS} sufficiently thoroughly that, under normal circumstances, if he/she takes him/herself to believe p_{AS}, to disbelieve p_{AS}, or to suspend judgment, then he/she *knows* that he/she does so believes, disbelieves, or suspends judgment.

Replacing *D2′* with *D2″* has no significant effect on the puzzle or the constraints on its correct solution.

Furthermore, the notion of reflection defined in *D2′* does not guarantee that if the reflective speaker takes him/herself under normal circumstances *not* to believe a proposition, he/she so takes him/herself correctly. Such an additional requirement would be excessive. In particular, I shall argue, although he is reflective, Pierre is very much mistaken *about himself* when under normal circumstances he continues to refrain in all sincerity from assenting to ‘Londres n’est pas jolie’. Though he does not realize it, he arguably believes exactly what he takes that sentence to express. Taking oneself not to believe is significantly different in this respect from taking oneself to believe. One can normally determine whether one believes a proposition p through careful consideration of the issue of whether *p* or ~*p*, and deciding between them. Deciding in favor of *p* in such circumstances is a way of believing *p*. Opting instead for ~*p* is a way of disbelieving *p*. As noted earlier, disbelieving is a kind of believing. It is not ipso facto a way of *not believing p*, of failing to believe *p*. Even failing to decide between *p* and ~*p* is not itself a way of failing to believe *p*. Deciding is not a way of not believing, and neither is failing to choose. Careful consideration of whether one believes provides some likelihood of being correct if one concludes that one does not believe. But it provides a *guarantee* of being correct if one concludes that one does believe. (See note 7.) In the case of *H4* the stipulation concerns the quality and character of Pierre's consideration: it is sufficiently thorough – sufficiently self-aware, truth-guided, thoughtful, careful, probing, dispassionate, unbiased, and so on – that if his circumstances are normal – if he is not under a hypnotic spell, not under the influence of hallucinogenic drugs, not manipulated by a Cartesian demon, and so on – and if he concludes that he really is of a certain opinion, this is appropriately precisely because he is in fact of that opinion. This may be taken to be every bit as *true by hypothesis* as any of *H1*–*H7*.

The notion of one act, event, or state of affairs appropriately occasioning another stands in need of clarification. Precisely what this amounts to does not affect the central issue. Pierre does sincerely assent, and he therefore has contradictory beliefs. Still, it is well to inquire into the relationship between verbal assent and belief. As I have argued at some length elsewhere, underlying the weaker disquotational schemata $D_{English}$, D_{French}, and so on, is the very nature of belief itself. Belief of a proposition is a favorable cognitive attitude. Embracing a proposition by believing it – as opposed to mere wishing or hoping – is, fundamentally, a kind of assenting. Belief is not mere outward, verbal assent to a sentence, however; more directly, it is a kind of inward, cognitive assent to the proposition itself.[14] This suggests a deeper definition for 'sincere assent':

> *D3''*: Agent *A sincerely assents to* sentence $S =_{def} A$ assents verbally to *S*; furthermore *A*'s verbal assent to *S* is appropriately an outward manifestation of *A*'s *cognitive* assent to p_{AS}, and therewith of his/her belief thereof, where p_{AS} is the very proposition he/she therewith takes *S* to express.

This alternative definition is significantly more illuminating than *D3*, especially in regard to understanding the legitimacy of $D_{English}$ and D_{French}.

[14] *Frege's Puzzle* (Atascadero, Calif.: Ridgeview, 1986, 1991), at pp. 80, 103–5, and passim.

Finally, the proposed definition for 'reticent' is straightforward:

D4: Agent *A* is *reticent* (to reveal his/her attitudes) *with respect to* proposition $p =_{def} A$ is not strongly disposed, or else is counter-disposed, to reveal (through assent, dissent, or abstention in response to queries) that he/she believes *p*, that he/she disbelieves *p*, or that he/she suspends judgment.[15]

It emerges from the proposed definitions, as well as from their potential replacements, that the stipulative hypotheses do invoke propositions – at least implicitly if not explicitly. For example, qua normal French speaker, Pierre stands in a specific relation to certain propositions. He interprets 'Londres est jolie' to express that London is pretty. What Pierre therewith takes the sentence to express – that London is pretty – is, even if Pierre does not recognize it, nothing more nor less than a proposition.[16]

III

The following important principle follows logically from definitions *D1*–*D3*. (As the reader can readily verify, it equally follows from *D1*, *D2′*, and *D3′*, and from *D1*, *D2*, and *D3″*.) The principle may be regarded as therefore analytic, on relevant interpretations of 'normal speaker', 'reflect', and 'sincere assent'.

B: For every commonplace sentence *S* of any language *L*, if a normal speaker of *L*, on reflection and under normal circumstances, sincerely assents to *S*, then he/she believes the proposition *S* expresses in *L*.

Put another way, substitution within *B* of the definitions of 'normal speaker', 'reflect', and 'sincere assent', results in a classical logical truth. This analytic truth can be employed in lieu of the disquotational schemata to generate the first version of Kripke's puzzle. Principle *B* together with *H1*, *H4*, and *H6*, and the further observation that French is a language and 'Londres est jolie' a commonplace French sentence, are

[15] Kripke: "The qualification about reticence is meant to take account of the fact a speaker may fail to avow his beliefs because of shyness, a desire for secrecy, to avoid offense, etc.… Maybe again the formulation needs further tightening, but the intent is clear" (p. 1015).

[16] Arguably the solution Kripke favors entails a rejection of *H1*, not as false but as untrue, on the ground that the phrase 'the proposition expressed in French by 'Londres est jolie'' is not well defined (is *improper*). I regard the rejection of *H1* on this ground as ill-motivated and excessively implausible. A similar situation obtains in connection with other hypotheses.

already sufficient to yield the result that Pierre believes the proposition expressed in French by 'Londres est jolie'. (See note 16.)

Affirmative answers to *Q1* and *Q2*, and therewith a hard constraint on any solution to Kripke's puzzle, are obtained as follows, using the proposed definitions *D1–D3* in place of the weaker disquotational schemata:

R3

a:	*D1–D3*	⊢	*B*.
b:	*B, H1, H4, H6, H8*	⊢	Pierre believes that London is pretty.
c:	*B, H2, H4, H7, H8*	⊢	Pierre disbelieves that London is pretty.
d:	*D1–D3, H1, H2, H4, H6–H8*	⊢	Pierre has contradictory beliefs.[17]

The analytic principle *B* does the primary work performed by the weaker disquotational schemata in Kripke's original formulation. Indeed, there is a clear sense in which *B*, which explicitly concerns propositions, underlies the weaker schemata. Though it is analytic, *B* might even be regarded as simply a more explicit rendering of those schemata. For example, $D_{English}$ and D_{French}, with substituends for 'φ' restricted to commonplace sentences, are derivable from *B* together with the following trivial schemata, respectively:

E: In English, 'φ' expresses (the proposition) that φ, and nothing else.
F: In French, 'φ_F' expresses (the proposition) that φ_E, and nothing else.

The schematic letter 'φ' is to be replaced by any suitable English sentence (containing no indexicals, etc.), 'φ_{F}' by any suitable French sentence, and 'φ_{E}' by its literal translation into English. This hard result may be formulated as follows:

R4

a:	*B, E*	⊢	$D_{English}$
b:	*B, F*	⊢	D_{French}*
c:	*D1–D3, E*	⊢	$D_{English}$
d:	*D1–D3, F*	⊢	D_{French}*

[17] As mentioned, these results, as well as the results to follow, are preserved if definition *D3* is replaced with *D3″*, or if *D2* is replaced with either *D2′* or *D2″* and *D3* is simultaneously replaced with *D3′*.

Instances of the schemata *E* and *F* are, strictly speaking, not themselves analytic. One does not know simply by virtue of one's knowledge of English that 'Londres est jolie' expresses in French that London is pretty. The instances of *E* and *F* are not themselves trivial. What is trivial is something meta-meta-theoretic: that every suitable instance of those schemata is true. By the same token, then, insofar as *B* is analytic it is trivial that every suitable instance of $D_{English}$ is true without exception, and similarly for D_{French}, $D_{Italian}$, and so on. This result supports $D_{French}{}'$ and $D_{English}{}'$, and therewith (given the appropriate stipulative hypotheses) the conclusion that Pierre indeed has contradictory beliefs. This result thus yields the same constraint on any solution to Kripke's puzzle.

As we have seen, Kripke objects to any such solution. His objection evidently makes use of a further hypothesis, one that is clearly more speculative than the purely stipulative hypotheses *H1–H7* and the trivially true *H8* and *H9*, to wit,

H10: If *H3*, then Pierre does not have contradictory beliefs.

As Kripke undoubtedly recognizes,

R5a: $\mathrm{H1–H4,\ H6–H10} \vdash \neg(\mathrm{D}_{\mathrm{French}}{}^{*} \wedge \mathrm{D}_{\mathrm{English}})$.

By insisting on *H10* in addition to the stipulative hypotheses, Kripke is committed to denying – erroneously if the foregoing is correct – that every instance of the weaker disquotational schemata is true. In particular he must reject as untrue (even if they are not false) the conjunction of $D_{French}{}'$ together with $\mathrm{D}_{English}{}'$.

On the other hand, as we have seen in *R3*, the proposed definitions *D1–D3* together with the stipulative hypotheses *H1*, *H2*, *H4*, and *H6–H9* yield the result that, for better or for worse, Pierre has contradictory beliefs. This yields an additional hard result, and with it an additional constraint on any solution to the puzzle:

R5b: $\mathrm{D1–D3,\ H1–H4,\ H6–H9} \vdash {\sim}\mathrm{H10}$.

This result discredits Kripke's objection. As already noted, hypothesis *H10* is speculative. Certainly it is more speculative than any of the hypotheses enumerated in *R5b*, each of which is either stipulated to be true by hypothesis or trivially true. Furthermore each of *D1–D3*, qua definition, is analytically true. Thus, true premises entail *H10*'s falsity. For better or for worse, *H10* is untenable. This result does not in itself solve Kripke's puzzle. A complete solution must acknowledge that Pierre has contradictory beliefs, and will also provide some account of how it happens that a rational agent in Pierre's situation excusably harbors contradictions.[18]

IV

Though the proposed definitions together with schemata *E* and *F* entail the weaker disquotational schemata, they do not also entail any of the strengthened disquotational schemata. In particular,

R6: $D1\text{–}D4, E \quad \nvDash \quad SD_{English}$

There are interpretations (models) on which *D1–D4* together with *E* are verified but $SD_{English}$ fails. As we shall see, one such interpretation is precisely the understanding of standard meta-English on which each of the definitions *D1–D4* provides the interpretations for 'normal speaker', and so on.

Although the strengthened disquotational schema cannot be deemed analytic or trivial, there are plausible, speculative hypotheses that support that schema. The most natural such speculative hypothesis is the following:

H11: If a speaker takes a sentence *S* to express a proposition *p*, believes the very proposition *p*, is disposed to reveal verbally that he/she believes *p*, and is not also counterdisposed, then under normal

[18] Kripke presents a fourth version of the puzzle (see note 9) on which *H6* is replaced with the following:

> ***H6′***: Pierre sincerely assents to 'Si New York est jolie, Londres est jolie aussi' and is not at all disposed to assent to 'Ce n'est pas que si New York est jolie, Londres est jolie aussi'. (p. 1022)

Correspondingly, *H4* is generalized and *Q1* is replaced with *Q1′*: 'Does Pierre believe that if New York is pretty then so is London?'. Pierre's inability to infer legitimately that New York is not pretty is evidently incompatible, given *H3*, with affirmative answers to both *Q1′* and *Q2*. A complete solution to this puzzle will answer both questions affirmatively and also provide an explanation of why Pierre's rationality does not enable him in this case to draw a simple modus tollens inference. Cf. my *Frege's Puzzle*, at pp. 103–18, 129–32; and "Illogical Belief," in J. Tomberlin, ed., *Philosophical Perspectives, 3: Philosophy of Mind and Action Theory* (Atascadero, Calif.: Ridgeview, 1989), pp. 243–85, reprinted in M. Davidson, ed., *On Sense and Direct Reference* (Boston: McGraw Hill, 2007), pp. 1037–67, and in *Content, Cognition, and Communication*, pp. 193–223.

circumstances he/she will be (more or less equally strongly) disposed to assent to *S*.

We have the following result:

R7:	*D1–D4, E, H11*	⊢	$SD_{English}$

I submit that the strengthened version of Kripke's puzzle, which employs the strengthened disquotational schemata, derives the bulk of its force from the plausibility of this hypothesis *H11* (or from that of similar speculative hypotheses).[19]

By the same token, *H11* is as untenable as *H10*. This follows from the preceding result:

R8

a:	D_{French}*, *H1, H2, H4–H9*	⊢	$SD_{English}$
b:	*D1-D4, H1, H2, H4–H9*	⊢	*~H11*

That is, the weaker disquotational schemata together with the listed stipulative hypotheses entail that not all instances of the strengthened disquotational schema is true. Insofar as *H11* is more speculative than the premises enumerated in *R8b*, this refutes *H11*.The stronger version of the puzzle is virtually an ironclad proof that $SD_{English}'$ is not true.

The latest result also yields a further constraint on any solution to Kripke's puzzle. The correct solution to Kripke's puzzle upholds the weaker disquotational schemata, providing affirmative answers to *Q1* and *Q2*, while rejecting the strengthened disquotational schema. A complete solution must also provide an explanation of how *H11* fails in cases like Pierre's. (See note 18.)

[19] The derivation of SD_{English} from *H11*, *D1–D4*, and *E* involves construing *D1* in such a way that for any commonplace English sentence *S* that univocally expresses only one proposition (with respect to a context), that same proposition is the only thing that a normal English speaker takes *S* to express (with respect to the context in question).