What Moore's Paradox Is About

CLAUDIO DE ALMEIDA
PUCRS (BRAZIL)

On the basis of arguments showing that none of the most influential analyses of Moore's paradox yields a successful resolution of the problem, a new analysis of it is offered. It is argued that, in attempting to render verdicts of either inconsistency or self-contradiction or self-refutation, those analyses have all failed to satisfactorily explain why a Moore-paradoxical proposition is such that it cannot be rationally believed. According to the proposed solution put forward here, a Moore-paradoxical proposition is one for which the believer can have no non-overridden evidence. The arguments for this claim make use of some of Peter Klein's views on epistemic defeasibility. It is further suggested that this proposal may have important meta-epistemological implications.

1. To see what it is about

Moore's paradox is solved here, I believe.[1] The natural first step in an attempt to show that my proposed solution is the only satisfactory account of so-called *Moorean absurdity* available—my task in what follows—is to supply the reader with the ground rules to which any serious discussion of the paradox must conform. Inspection of the literature on the issue reveals that all of the following constraints have been regarded (though not always explicitly) as constitutive of the challenge posed by the need to explain the peculiar form of absurdity we call *Moorean*.

First, we want to understand what makes (the objects of) certain *beliefs* Moore-paradoxical.[2] Given this aim, the prevailing attitude is the one according to which an adequate explanation of the nature of Moorean absurdity is, first and foremost, an explanation of the oddity of (the objects of) certain *beliefs*; and it is thought that this will bring, as a bonus, an explanation of Moore-paradoxical *assertion* "via the principle", in Shoemaker's (1996, p. 76) words, "that what can be (coherently) believed constrains what can be (coherently) asserted" without the converse being the case.[3]

---

[1] For a particularly good brief description of the problem, see Roy Sorensen 1992. Moore's own discussion of it is found in his 1942, 1944, and, best of all, 1993.

[2] As far as I know, use of the expression 'Moore-paradoxical' originates either in Sydney Shoemaker 1988 or in Laurence Goldstein 1988.

[3] Essentially the same considerations that would suffice to disqualify attempts to explain Moorean absurdity which are narrowly based on a notion of absurd *assertion* could be

Second, we seek an explicit definition of the predicate 'Moore-paradoxical proposition'.[4] This is not, of course, to say that those analyses of the problem which do not produce an explicit definition of the predicate do not merit attention. But there should be no doubt that we cannot be as clear about the nature of Moorean absurdity as we hope to be as long as we fail to produce such a definition. Mine will be offered in section 5 below.

Third, any acceptable analysis must offer an explanation of the fact that a Moore-paradoxical proposition seems to be both contingent and *contradiction-like*. That is to say, we are required to expose that property of Moore-paradoxical propositions in virtue of which, in believing some such apparently contingent proposition, "one", as Sorensen (1988, p. 16) puts it, "*in some sense*, contradicts oneself" (emphasis mine).[5]

Fourth, in deeming a Moorean believer *irrational*, which is what we all want to do, we have to make sure that our criticism of him relies solely on principles the violation of which can safely be considered sufficient to warrant pinning the label 'irrational' on the violator. In other words, we expect our criticism of a Moorean believer to yield a partial definition of *rationality* by uncovering some principle(s) the observance of which we would be willing to regard as a necessary condition of one's being rational.[6] I shall henceforth refer to this constraint as "Condition R".

---

brought to bear against the suggestion that the key to understanding Moorean absurdity might be provided by a notion of "pragmatic contradiction". For a suggestion along these lines (in which, to be sure, no overt claims about Moore's paradox are made), see Richard Foley 1995. Also, see my 1999.

[4] In what follows, I adhere to the practice of speaking about *propositions* even when I ostensibly refer to *sentences* (which presumably express those propositions), thus avoiding circumlocution. Also, while it is acknowledged that problems for the identification of propositions *in general* may arise from considerations having to do with indexicality, it is understood that proposition-talk is harmless enough in this context. This is standard practice in discussions of Moore's paradox, and I don't see that it is a bad one.

[5] Surprisingly to me, Sorensen (1988, p. 40) describes belief in a contradiction as a "more grievous offence" than belief in a Moore-paradoxical proposition, which he describes as a "lesser misdemeanour". But I take it that our job should be one of showing that belief in a Moore-paradoxical proposition is *just as bad* as belief in a contradiction.

[6] In my use of the term here, 'rationality' is short for 'epistemic rationality'—the term we would use, provided certain conditions were satisfied, to describe the constitution or maintenance of a doxastic system in pursuit of the "purely epistemic goal", in Foley's words, "of now believing those propositions that are true and now not believing those propositions that are false". See Foley 1993, chapter 1, for a discussion of both this and the broader notion of rationality according to which "[t]here is no single perspective that is adequate for understanding the entire range of our judgments of rationality"—for Foley, not even the more restricted range of our judgments of epistemic rationality. (Although I find Foley's conclusion about epistemic rationality unduly pessimistic, I cannot pursue this matter here.) I further assume that epistemic justification is at least a necessary condition of epistemic rationality. My use of the term 'epistemic justification' is derived from—but is not identical with—Peter Klein's 1981 use of it. In his terminology, justification is to be equated with "groundedness" plus the absence of "external

Fifth and last, we expect a solution to the problem to be *complete*—that is, to explain the oddity of all instances of Moorean absurdity. As far as I know, Sorensen has single-handedly promoted awareness of the need for the satisfaction of some completeness condition; but it is not clear to me whether he sees that we may not be in a position to enforce such a condition. The problem here arises from the fact that we do not seem to have a clear intuitive grasp of Moorean absurdity. We usually identify it with the propositional forms exemplified by obvious instances of it, such as

(1) It is raining, but I believe it isn't

and

(2) It is raining, but I don't believe it is,

which are instance of the forms p&B~p and p&~Bp, respectively—if we read, as we shall do henceforth, 'Bp' as 'I believe that p'. However, despite the fact that our intuitions do not seem to yield even a rough and ready syntactic or semantic criterion of Moorean absurdity, Sorensen clearly seems right when he claims that some propositions ought, on reflection, to be considered Moore-paradoxical either because they imply instances of those readily identifiable Moore-paradoxical forms or because *some* (but not all) believers cannot rationally have them as objects of belief. Thus, he draws our attention to

(3) Although you do not agree with me about anything, you are always right,

(4) The atheism of my mother's nieceless brother's only nephew angers God,

and

(5) It is raining, but *you* don't believe it is.

Assuming with him, as we certainly should, that "disagreement implies lack of shared belief" (1988, p. 47), (3) implies "All of your beliefs are true, but I don't believe they are", which is a more obvious case of Moorean absurdity than (3). Likewise, (4), as he points out (1988, pp. 28–29), is Moore-paradoxical, since it implies "My atheism angers God", and it is quite reasonable to assume, with him, that this in turn implies "God exists, but I believe that God does not exist".

---

overriders" (not to be confused with what he calls "defeaters"). Except for what I deem a questionable reliability requirement (see my 1998), his notion of justification coincides with what Keith Lehrer (1990) calls "personal justification".

Sorensen is again right in thinking that (5) is Moore-paradoxical *to some degree*, since believing it *sometimes* (but not always) commits the believer to believing (2). Anyone can believe (5) without absurdity *unless* the believer is himself the object designated by the pronoun 'you' in (5). Let us follow Sorensen here and say that a proposition is Moore-paradoxical the way (5) is only if it is Moore-paradoxical for the believer *to whom it is addressed*. Thus, if I am the *addressee* of (5), that proposition is Moore-paradoxical for me alone, because, in believing (5), I become committed to believing (2).[7] We really have, therefore, to distinguish two *types* of Moorean absurdity in order to be sensitive to the fact that some propositions are *always*—whereas others are only *sometimes*—Moore-paradoxical for the believer, as illustrated by the contrast between (5) and all of the other Moore-paradoxical propositions we have so far considered.

Sorensen remains right when he claims that both "It is raining, but I doubt that it is raining" and "It is raining, and I guess that it is raining" are Moore-paradoxical. If doubting whether p is the same as neither believing nor disbelieving (denying) p, and if guessing that p implies lack of belief in p, then both of the above propositions imply (2) by describing propositional attitudes which are contrary to belief.

Further, even though he does not explicitly consider either

(6) It is raining, but I'm not justified in believing that it is

or

(7) It is raining, but I'm justified in believing that it is not,

I trust he would have conceded that both are Moore-paradoxical. Despite the fact that neither (6) nor (7) describes any propositional attitude whatsoever—since having justification for believing p does not imply that p is believed—it seems clear that, once we assume that belief distributes over conjunctions (that is, S believes that p&q only if S believes that p and S believes that q), what is so often done and will be done here, it would seem that the believer both believes that it is raining and believes that this is not the epistemically right thing for him to do in (6) and believes the contradictory of what he believes himself to be epistemically obliged to believe in (7), which ought to be regarded as an offense of the Moore-paradoxical kind.[8]

---

[7] In introducing this notion of the *addressee* of a proposition, I do not feel the need to be any more formal than Sorensen has been.

[8] Even though she's not explicitly concerned with Moore's paradox, Doris Olin 1983 (p. 229) seems to have considered it *obvious* that "[i]t can never be reasonable to believe a proposition of the form '*p* and I am not now justified in believing *p*'". She argues for her claim as follows:

Now, is Sorensen right in claiming that both

(8) I have no beliefs now

and

(9) I don't exist

are to be regarded as Moore-paradoxical? Clearly, in considering the way in which (8) and (9) strike us as being absurd, we are way past the point where a simple appeal to our intuitions concerning Moorean absurdity can help us determine whether a given analysis of the problem has satisfied a completeness condition. Appearances won't matter much, though, where they clash with arguments which we perceive to be sound. Thus, I shall be arguing that Sorensen *is* right in taking both (8) and (9) to be Moore-paradoxical (even though, if I am right, his own analysis of Moorean absurdity fails).

Finally, the foregoing seems to indicate that a sober approach to a completeness condition would require that passing the test posed by (1), (2), (5), (6), and (7) be taken as a *sine qua non* condition for any proposed solution to Moore's paradox—we might call it a "minimal completeness condition". Even so, in what follows, I am willing to waive the completeness condition when briefly discussing some important recent proposals, since, with the possible exception of Sorensen's, they are all oblivious to (5), (6) and (7), and I work on the assumption that we could conceivably learn from an attempt to explain Moorean absurdity that is not ostensibly concerned with satisfying the completeness condition. My own proposal, at all events, is expected to pass the test posed by all of the Moore-paradoxical propositions we have identified. And more: on my account of the matter, we also get an explanation of the oddity of (8) and (9), and, with this, a reason to believe that Moorean absurdity is an even more important epistemic phenomenon than we might expect when first encountering it.

[I]f a person *A* is justified in believing a proposition, then he is not (epistemically) blameworthy for believing it. But if *A* is justified in believing that he is not justified in believing *p*, then he would be at fault in believing *p*. Hence, if *A* is justified in believing that he is not justified in believing *p*, then he is *not* justified in believing *p*.

As it stands, however, her argument seems vulnerable to the following objection. As long as justified belief can be false, A surely can be justified in believing that he is not justified in believing p and yet be justified in believing p (since he can have justification for the false belief that he is not justified in believing p). But if, as I believe, Olin's intuition leads us in the right direction, there ought to be a sound argument to the effect that one *cannot be justified in believing* a proposition of the form instantiated by (6). In section 5 below, I put forward an argument for this claim.

2. It is not about self-contradiction

Arguably, Wittgenstein did more to promote awareness of the philosophical significance of Moorean absurdity than Moore himself. The lasting influence of his remarks on the problem bears witness to this.[9] But, when we turn to the recent analyses of Moorean absurdity inspired by those remarks, we see that the popularity with which the Wittgensteinian approach to the problem has met far exceeds its plausibility.

Consider the claims made by Kent Linville and Merrill Ring (1991), whose "overall strategy... is designed to show that 'I believe it is raining but it's not', for example, is absurd because it consists of two contradictory assertions about the weather" (p. 296). According to them, "'I believe that p' announces a verdict about the truth-value of p, not the condition of the person rendering that verdict; and others can say of me that I seem to believe because their words—'He seems to believe that p' (said of me)—do form a hypothesis about me, not a verdict about p" (p. 303).[10] In spite of the differences in their analyses of the paradox, both Arthur Collins (1996) and Jane Heal (1994) close ranks behind Linville and Ring's unmistakably Wittgensteinian claim that "'I believe' and 'He believes' are different instruments which perform functions as distinct as that made familiar by John Austin between 'I promise' and 'He promises'" (Linville and Ring, p. 303)—these functions being, of course, indicated by the alleged fact that, in saying 'I promise that p', I *do* the promising, whereas I *only describe* a promise when I say "He promises that p". This certainly is the crucial contention of a Wittgensteinian approach to Moore's paradox. It indicates how the Wittgensteinian analyst hopes to substantiate the claim that belief in a Moore-paradoxical proposition is an act of self-contradiction. In a rhetorical move that is not unlike slapping the reader in the face, in hopes of bringing him out of his philosophical stupor, the Wittgensteinian expects to see it acknowledged that our use of the expression 'I believe' in our ordinary linguistic transactions merely indicates our commitment to whatever it is that we claim to believe. In Collins' words, "the two linguistic devices, assertion of p and assertion of 'I believe that p', are alternative means for making the same conversational move" (p. 316). So, as Heal claims, when somebody says 'I believe that p, but not-p', "[i]n effect what has been said is 'p but not-p'" (pp. 20–21).

A very evident problem with this account of the matter is that we *do* feel the need to describe our propositional attitudes in certain very ordinary contexts, as when we say "I believe the butler did it, but I can't say I know it", or "This is how we stand: I believe you're innocent; she doesn't", or "I believe he can make it, because the evidence that he can is overwhelming", or

---

9   His most important remarks on the problem are in his 1953, pp. 190–92, and in his 1980, pp. 90–96. For the relevant history, see Baldwin 1990, pp. 226–32, and Gombay 1988.
10 This is echoed in Collins (1996, p.315).

"Rest assured I believe you're not one of them". On present showing, however, the Wittgensteinian analyst would be either hopelessly wrong about these uses of 'I believe' or hard-pressed to explain the connection between these (deviant?) uses of 'I believe' and the one in which, allegedly, the speaker *merely* assigns T to p in saying "I believe that p".

Granted, while this objection seems to have been ignored by some among Wittgensteinians, it will not, by itself, stump Heal, whose proposed solution to Moore's paradox is encapsulated in the following passage (Heal, pp. 22–23).

[W]e have in "I believe that p" an utterance which is, at one and the same time, a member of two different classes. On the one hand it is a self-description of me as a believer and as such it has all the possibilities of grammatical transformation...On the other hand it is an expression of belief that p, an alternative way of voicing out what could also be voiced out as "p". When we sense the contradiction in the Moorean utterance we hear "I believe that p" in this second role. When we become puzzled about why the utterance is contradictory we hear it in the first role...When we contemplate someone having this thought ["I believe that p"] we take him or her to have a representation which is at the same time about the self and about the world. It is both a belief that he or she believes that p and a belief that p.

As it turns out, however, the Wittgensteinian will find no refuge in Heal's claim to the effect that 'I believe' has the two "roles" of *asserting* whatever is said to be believed and of *describing* the relevant propositional attitude, and that such roles are somehow simultaneously played in an assertion of (or belief in what is expressed by) 'I believe that p'. For we now could simply choose to call "Moore's paradox" the absurdity exemplified by that facet of (1) or (2) that is absurd but is not regarded by Heal herself as a contradiction! *That* absurdity has been left untouched by her! In other words, having acknowledged that there is a purely descriptive role for an assertion of 'I believe that p', Heal cannot explain what is puzzling about that descriptive role.[11]

It is also noteworthy that the Wittgensteinian fails to satisfactorily account for the negations of (1) and (2). If (2) is a contradiction, its negation

---

[11] Notice that we can still consistently accept, with Norman Malcolm (1995, p.197), the much-cherished idea according to which our everyday use of 'I believe that p' "commonly amounts to...a *tentative* or *hesitant* assertion of 'p'". (Alternatively, I suppose it may be possible to maintain that such a use of 'I believe' only indicates an *inclination* towards the assertion of p.) However, it seems to me remarkable that the Wittgensteinian analyst would still try to stand his ground when he acknowledges that 'I believe' may, after all, be put to a puzzling descriptive use but simply cannot explain what is wrong with that descriptive use of the verb 'to believe'. Thus, Malcolm (p. 205), after offering the usual Wittgensteinian considerations, goes on to concede that

[I]t is true that in certain situations someone who says 'I believe such and such' *is* talking about himself. I do not wish to give the impression that Moore's paradox arises *only* when the speaker is talking about the such and such, and not when he is talking about himself.

must be a tautology. Yet, as Sorensen (1988, p. 15) notes, "'Either it is not raining or I believe it is raining' fails to be a necessary truth even in the mouths of our most marvelous meteorologists". The negation of (1) is not a tautology either, since it is logically equivalent to the ludicrous 'I believe it's not raining only if it's not'.

Granted, it is open to the Wittgensteinian analyst to claim that this last objection begs the question.[12] It may be thought that *denouncing* the classical notions of negation and logical truth and falsehood is the crucial element of the Wittgensteinian enterprise that the objection ignores. But, to my mind, this line of response is a dead end. Never mind the fact that the Wittgensteinian analyst has failed to provide us with the non-classical counterparts of those loathed classical notions. (He will tell you not to turn to him for such counterparts, since a *compositional* account of meaning is precisely what he opposes.[13]) He has now cornered himself into merely pleading for the acknowledgement of the *contradiction-like* quality of (1) and (2); but the rhetoric of such a plea is driven by an appeal to the only notion of contradiction we are familiar with, *our* classical notion. So, the motivation for brandishing the label 'contradiction' at just about *any* piece of discourse which displays *some* contradiction-like quality falls flat as long as a less dramatic explanatory move is available to us.[14]

---

[12]   That it may be thus seen was brought to my attention by Paulo Faria.
[13]   See Goldstein 1988, pp. 70–71.
[14]   To my mind, Goldstein's interpretation of the relevant Wittgensteinian views is as compelling as any when he claims that "the solutions Wittgenstein proffers for Moore's paradox and for the logico-semantical paradoxes are identical" (Goldstein 1988, p. 49). Even though a detailed examination of such views far exceeds the scope of this paper, I feel I should voice my profound scepticism about their tenability. As far as I've been able to determine, on Goldstein's account of the matter, the relevant cluster of *late* Wittgensteinian views is held together by the claims embodied in the following readily identifiable piece of Wittgensteinian catechism.

> Our grasp of the rules for 'and' and 'not' is manifested *inter alia* by our excluding contradictory assertions, except where these words are used in nonstandard senses, or where we can supply 'completing surroundings' for what would otherwise be a defective performance. *Refraining* from affirming 'not-p' when you believe that p (unless you want to lie, or other 'peculiar surroundings' prevail)...is part of what it is to correctly use 'not'...Nothing *makes* it a correct way to use that word—that's just what we call using 'not' correctly...

As already suggested in this passage, the Wittgensteinian rationale for a quick dissolution of some of the most mind-boggling paradoxes handed down to us by the philosophical imagination will crucially depend on there being nonliteral uses of tautologies and contradictions ("If there is no preservation of the 'literal meanings' of expressions in, for example, metaphor, euphemism and hyperbole, how likely is it that these meanings are preserved in such nonstandard utterances as candid assertions of contradictions?" [Goldstein 1988, p. 71]). Apparently oblivious to the fact that a compositional account of meaning should not automatically either be equated with "a Platonistic conception of meaning" [cf. Goldstein, op. cit.] or, what matters most here, be presumed incompatible with a satisfactory explanation of nonliteral discourse, the

3. It is not about self-refutation

According to John N. Williams (1994), (1) and (2) are absurd for different reasons. As regards (2), his proposed solution is based on the observation that belief in (2) is self-refuting (or self-falsifying). He rightly sees that, on the seemingly unproblematic assumption that belief distributes over conjunctions, anyone who believes an instance of the form p&~Bp must believe falsely, since the required belief that p falsifies the conjunctive belief by falsifying what is described by one of the conjuncts (~Bp). This is, in fact, the case: belief in (2) *is* self-refuting. But is this what the *Moorean* absurdity of (2) is about? I submit that it is not.

We have agreed that an acceptable explanation of Moorean absurdity calls for the satisfaction of our Condition R. Now, if what is essential to Williams' explanation is the claim according to which, in believing (2), one *must have false belief*, his explanation is too broad to satisfy Condition R, since a sure-fire way of holding a false belief is to believe a necessary falsehood. Yet, it should be granted that one may rationally hold belief in a necessary falsehood. (One unhappy evening, in serious conversation, the very reliable Kleene asserts what, unbeknownst to me, is the negation of the Löwenheim-Skolem theorem, and I believe him. Furthermore, think of the metaphilosophical view implied by the opening remarks to Grice's 1961, according to which every philosophical claim is either necessarily true or necessarily false.[15] Those who espouse such a view are certainly right in thinking that, if it should turn out that they are wrong about some controversial philosophical issue, there may still be no basis to their being called "irrational".)[16]

Maybe this is not a charitable enough account of Williams' analysis of the Moorean absurdity of (2). Maybe, for him, self-refutation is not to be taken as lightly as I may have taken it: he may think that what makes one

---

Wittgensteinian analyst fancies that the way has been paved for his fundamental claim according to which *it is a mistake to think that contradictions are false*. But look closely at the claim with which Goldstein concludes the passage quoted above: "Hence to candidly affirm 'p and not-p' is to use 'not' *incorrectly* and this is quite different from using it correctly to make a necessarily false assertion" (Goldstein, p. 70). Does the Wittgensteinian believe that every assertion of a necessarily false proposition is either a (senseless) misuse of language or a piece of nonliteral discourse? Indeed, on Wittgensteinian premises, is it at all conceivable that one may *correctly* use language to make a necessarily false *assertion*? It seems clear to me that no articulate Wittgensteinian answer to these questions is forthcoming.

15 Bertrand Kolecza drew my attention to this passage in Grice.
16 After noting, on *slightly* different grounds, that (2) cannot be truly believed, Shoemaker (1996, pp. 77) asks whether it can be *falsely* believed and answers as follows: "Well, since it is demonstrably true that [(2)] could not be truly believed, anyone who believed it would be in a position to see that his belief had to be false." As I understand, the same can be said about belief in a necessary falsehood; yet, such a belief could be rationally held. Shoemaker seems to imply—and wrongly so—that *being in a position to see* that one believes falsely should suffice to justify our charging the believer with irrationality.

who believes (2) irrational is not *just* the fact that he must hold a false belief, but the fact that the believer *falsifies his own belief*. Trouble is, this account of the matter does not give us all it takes to satisfy our Condition R. It would also be necessary to establish that, while believing (2), the believer must be aware of the fact that he falsifies his own belief. Surely, unless it is obvious to me that I falsify my own belief, I cannot be deemed irrational. As a rule, one's rationality is not threatened by the fact that one is mistaken about what one has done. (I may infer, from my false but justified belief that eating cheese is always conducive to weight loss, that I'm losing weight while eating an amount of cheese that is, in fact, incompatible with loss of weight.)

As I see it, however, the two individually necessary and jointly sufficient conditions of my being aware of the fact that I have refuted my own belief, when believing an instance of p&~Bp, will hardly be taken as necessary conditions of rationality. First, Williams' account of the Moorean absurdity of (2) depends on showing that believing is unlike actions of other types insofar as it is *self-intimating*—that is, it is such that, necessarily, S believes that p only if S (perhaps dispositionally) believes that he believes that p, a most controversial claim. Second, the believer must present himself with something like Williams' own argument for the absurdity of (2)!

4. It is not about inconsistency

The other half of Williams' account of Moorean absurdity relies on the fact that anyone who believes a *true* instance of p&B~p holds contradictory beliefs, thus being the bearer of an inconsistent belief set. Indeed, if I believe (1) and my belief is true, then I have both the belief that it is raining (because I believe that p) and the belief that it is not raining (because, the conjunction being true, B~p is true). Therefore, if I believe (1), either my belief is false or I am inconsistent (for holding contradictory beliefs). Therefore, if I believe (1), I must have a false belief (since inconsistency is a guarantee of the presence of a false belief in one's belief set). Is this why (1) is Moore-paradoxical?

Apparently not. Williams neglects to tell us why (1) cannot benefit from a comparison with the belief giving rise to the Preface paradox. It seems one can *justifiably* believe the following: 'At least one of my *other* beliefs is false'. (Call it "the prefatory belief".)[17] If you have such a belief, you must have a false belief: either the prefatory belief is false or you are inconsistent. To be sure, there *is* a potentially significant difference between the two cases, a difference stemming from the distinction between *weak* and *strong*

---

[17] Contrary to what John Etchemendy 1995 claims, the prefatory belief need not be a contingent liar. See Sorensen 1988, pp. 23–24.

inconsistency.[18] You are strongly inconsistent if and only if you believe contradictories. All other cases of inconsistency—all other cases in which your beliefs cannot all have the same truth-value—are cases of weak inconsistency. Will this vindicate Williams' claim about the Moorean absurdity of (1)?

Clearly, the answer to this must be: Not just by itself. The distinction between weak and strong inconsistency *might* be the first step in an attempt to satisfy Condition R. But, unless it is shown why strong inconsistency is not a price the believer may acceptably pay for believing a proposition that seems epistemically justified for him (on both evidentialist and reliabilist grounds, as we shall now see), we are left without a compelling case against believing (1). And, it certainly *seems* obvious that one could have (non-overridden) evidence for believing an instance of the form p&B~p. My psychoanalyst, whom I *know* to be an extremely reliable source of information about both my beliefs and the facts of my life, tells me: "Your father loves you, but you believe he doesn't". I believe her—and rationally so, it would appear. Therefore, on present showing, I rationally believe 'My father loves me, but I believe he doesn't', and become strongly inconsistent in the process. We cannot *just assume*, as Williams does, that strong inconsistency is never a price to pay for rational doxastic behavior.[19]

Williams' case against belief in (1) seems to depend on insensitivity to the sting of the Preface paradox. If the Preface strikes you as a case of rationally sustained inconsistency, you have to demand an account of the distinction between acceptable and unacceptable forms of inconsistency. In the absence of such an account, you may find refuge in Sorensen's remark according to which, when rationality is in question, "consistency is just one desideratum amongst a set of competing desiderata"—"competition with other desiderata sometimes leads to permissible types of inconsistency" (1988, p. 23).[20]

---

[18] See Klein 1985. The distinction is also recognized by Sorensen 1988.

[19] As we shall see, in the proposed solution put forward in section 6 below, strong inconsistency is ruled out as a case of a more-encompassing phenomenon: incoherence. My proposal does not simply *assume*, as Williams' does, that strong inconsistency is always objectionable.

[20] While believing that inconsistency is essential to Moorean absurdity, Sorensen (1988) has managed to keep a distance from the kind of blanket condemnation of inconsistency suggested by Williams' analysis. But this comes at a price we most definitely would not want to pay. His definition of 'pure Moorean proposition', on which his explanation of Moorean absurdity rests, implies that one who believes either (1) or (2) is irrational because he is not "absolutely thorough", according to the following definition of *absolute thoroughness* (p. 28): "An agent is *absolutely thorough* at [time] *t* just in case his beliefs are deductively closed and distribute over material conditionals at *t*". But charging the believer with failing to live up to the ideal of logical omniscience does not bring us even close to satisfying Condition R. We certainly are in possession of a concept of rationality which is partly shaped by the belief that logical omniscience is, *in Sorensen's own words,*

5. What it is about

Moore's paradox, I submit, is essentially about a certain necessary condition of epistemic justification. This, of course, is not saying much. Moreover, it is saying something with which all those who have put forward the proposals discussed above would probably agree, since self-contradiction, self-refutation and inconsistency may all have an impact on one's ability to pursue the epistemic goal of constituting and maintaining a doxastic system with a view to believing only true propositions. But, as we have seen, chiefly through inattention to the constraints imposed by our Condition R, those proposals have failed to identify what it is about a Moore-paradoxical proposition that makes believing it a mental act that cannot be justified, thus rendering the doxastic agent vulnerable to the charge of irrationality. (Again, I assume that every necessary condition of believing with justification is a necessary condition of believing rationally.)

My first substantial claim here is that a Moore-paradoxical proposition is one for which the believer can have *no non-overridden evidence*.[21] What follows is an attempt to establish this claim.

My second substantial claim is that the identification of that property of justified beliefs that is missing in the case of Moore-paradoxical beliefs is at least greatly facilitated by use of the conceptual framework of the best *general* theory of evidential support that I'm familiar with, Klein's defeasibility theory of epistemic justification.[22] In his own words, his "defeasibility model of justification" is concerned with "the conditions governing the epistemic

---

"a waste": "[o]ur limited intellectual resources should [only]...be assigned to the more interesting consequences of our beliefs" (p. 36).

   For Shoemaker (1996), inconsistency is also at the heart of Moore's paradox. According to him, by putting into question the thesis that *rationally held* belief is self-intimating, the paradox turns out to be essentially about the nature of self-knowledge— "specifically, knowledge of one's own beliefs" (p. 92). Space limitations keep me from discussing his proposal in detail here. Suffice it to say that he wants to use the principle according to which belief distributes over conjunctions in cahoots with the claim that I rationally believe that p only if I (at least dispositionally) believe that I believe that p (self-intimation) to show that believing (2) makes me strongly inconsistent. It's a costly strategy, since it (a) fails to take us any further than Williams has, (b) adds the troublesome self-intimation claim to our conceptual baggage, and (c) does nothing to explain the oddity of (1), which, as Shoemaker himself admits, would, on his conceptual framework, require the use of an even more troublesome principle.

[21]   The impact of this claim depends on our assuming, as I do, that having (non-overridden) evidence is a necessary condition of epistemic justification. This assumption places my proposal squarely into the evidentialist camp as regards the dispute concerning the nature of epistemic justification. But, in section 6 below, I think I have hair-raising news for the reliabilist who might want to shrug at my evidentialist stance.

[22]   As we know, we have several accounts of epistemic defeasibility at our disposal. Klein's is, from my standpoint, the most resourceful of them. But comparing it to its competition or addressing certain criticisms that have been leveled against it is not my job here. At all events, I should like to mention that there is reason to think that Plantinga's 1996 attack on Klein's defeasibility theory is a failure. See my 1998.

acceptability of a set of propositions" (Klein 1985, pp. 111–12). I describe it as "a general theory of evidential support" because it assumes that the epistemic acceptability (that is, acceptability with a view to the acquisition of truths only) of a proposition for an agent S hinges on S's possession of *evidence* for the proposition and makes it its job to spell out the conditions on which evidence can be epistemically effective in providing support for that proposition in S's belief system.[23] It seems accurate to say that what characterizes this epistemological project is the attempt to specify the *systemic* conditions for evidential effectiveness. Much of the motivation for the defeasibility theory is provided by the need to show how the evidence one has for believing a given proposition may be defeated by counterevidence one does not possess (cf. Klein 1981 and 1996). This seems essential for a discussion of the Gettier problem, but need not concern us here. Fortunately, the conceptual backing that we need from Klein's theory in order to generate an adequate definition of 'Moore-paradoxical proposition' does not require an exposition of his theory that brings all of its complexity into view. So, I shall expose only those elements of the theory which are absolutely essential for our purposes.[24]

Let us begin by saying that, if there is enough evidence (whatever 'enough' may mean) for a proposition p in S's belief system and the evidence is not defeated by effective counterevidence, p is *warranted* for S. (To elucidate what 'counterevidence' is supposed to mean in this context, suffice it to say that a proposition e is counterevidence for p if and only if the addition of e to S's belief system sufficiently weakens S's epistemic right to believe that p either because e itself is an indication of the falsehood of p or because e is in the evidential ancestry of a proposition which is an indication of the falsehood of p. I expect what is meant by the terms 'effective counterevidence' and 'evidential ancestry' to be clear in what follows.) Thus understood, warrant is intended to be a justification-conferring relation between propositions.[25] We want to say that p is justified for S only if p is warranted for S.

---

[23] In my understanding of his 1996 paper, Klein in effect ceased to be an evidentialist and embraced reliabilism. This, in my opinion, was a dramatic and *ill-advised* move on his part. See my 1998. Here, in any case, his apparent change of heart won't matter, since his account of defeasibility—as we might expect of such accounts in general—is not obviously incompatible with reliabilism.

[24] My main sources in what follows are Klein 1985 and 1986. To see how the defeasibility theory of *justification* provides the basis for a defeasibility theory of *knowledge*, see Klein 1981.

[25] In Klein's own words, the appeal of his theory is expected to come partly from the fact that "one of the most important things which [can] be left unsaid is a characterization of what is required in order for a proposition to be warranted to a degree appropriate for justification" (1986, p. 265). The reader is also cautioned against confusing Klein's use of the term 'warrant' with Plantinga's 1993 influential use of the term.

Further, warrant for a proposition p, for an agent S, can always be traced back to S's (actually held, but not necessarily occurrent) beliefs, those beliefs which provide enough evidence for p in S's belief system. (Let us use 'eWp' to say that proposition e warrants p, and 'e[W]p' to say that e fails to warrant p.) Thus, we speak of *warrant paths* extending from the propositions that one believes to the propositions that one is entitled to believe given one's present stock of beliefs.[26]

Warrant is defeasible. But we are not presently concerned with all the varieties of epistemic defeasibility. For our purposes, it should suffice to recall that there are cases in which a warrant path is blocked by counterevidence residing in the very path it blocks—i.e., that there is what Klein calls "internal blocking". Consider the case in which eWqWp and it is not the case that eW~p. There could be a proposition r, not in S's belief system, which is warranted by e and is such that (r&q)[W]p (regardless of whether r by itself is evidence against p). In this case, the relevant piece of evidence against p, the blocker of the warrant path to p, namely e, comes from the evidential ancestry of p itself. Here, it is as if e were a double agent: it is instrumental in both providing and destroying the evidence for p. Since the evidence against p originates in the warrant path to p itself, we say that the warrant path to p is *internally blocked*.[27]

However, since warrant is not a transitive relation, the very proposition which originates the evidential path to p may itself be the evidence against p. This can be illustrated by one of Klein's own examples, the Clever Car Thief Case (cf. Klein 1981 and 1985). Suppose S believes all of the following:

s: Jones is a clever car thief and has stolen a car.

r: Clever car thieves behave exactly like legitimate car owners.

q: Jones behaves as if he owned the car he drives.

p: Jones owns the car he drives.

---

[26]   An important word of caution. Inattention to Klein's use of the familiar distinction between dispositional and occurrent beliefs (*not to be confused with a conscious/unconscious distinction*) may give the impression that his concept of warrant is one according to which only propositions can be justifiers. Thus, a referee for this journal criticizes me for adopting what he deems an obviously defective conception of warrant. In fact, the referee has failed to notice that Klein's talk of propositions as being the only justifiers is convenient *façon de parler*. It certainly is plausible to say that, if my experience X warrants (or prima facie justifies) belief in proposition p for me, I at least *dispositionally* believe that I have X and the relation between X and p can be expressed by saying that the proposition expressed by "I have X" warrants p for me. See Klein 1981, pp. 33 and 44–47. To be sure, the referee could justly complain that Klein has been a bit too casual in dealing with this kind of serious misunderstanding of his views.

[27]   "Evidence against p" is, of course, shorthand for "evidence for disbelieving p".

Here, although (s&r)Wq and qWp, (s&r)[W]p.[28] The path to p is internally blocked, that is, it is blocked by a proposition which is in the evidential ancestry of p. (We might as well say that there is no warrant path to p. But the metaphor of a *blocked* path helps us understand how paths get destroyed.)

But the effect of a blocker may itself be canceled: there may be a blocker for the path to the blocker. Thus, in the above case, if we added

t: Jones has been framed by the police

to S's belief system, conjoining the evidence leading to s with t blocks the path to s. The proposition s would then have to be excluded from S's belief system, which, in turn, would restore the warrant for p. The principle governing such an exclusion is what Klein calls "Rule of Revision"—the rule according to which, "[w]henever a belief is added to a belief system, any belief which would [block the path to] that belief must be subtracted [from the belief system]" (Klein 1986, p. 266).[29] We need some such rule to allow for belief revision while excluding objectionable forms of incoherence, particularly strong inconsistency.[30]

In Klein's jargon, a *clear* warrant path is one which is either unblocked or, if it is blocked, the effect of the blockers is itself nullified by blockers which restore the warrant for the proposition. A path in which this second possibility obtains is one which is *ultimately unblocked* (cf. Klein 1985, p. 115). Let us then make an addition to Klein's terminology here to make it relevant to the only case of blocking with which we are concerned, internal blocking, and say what it is for a path to be *internally clear*.

A warrant path is *internally clear* if and only if either it is not internally blocked or it is ultimately internally unblocked

---

[28] A referee for this journal suggests that it must be wrong to say that q by itself warrants p. The referee thinks that it is the conjunction of q with something like the background belief that all those who engage in perfect car-owning behavior probably own the cars they drive that warrants p. But, since we are not interested in explaining *why* this is a world in which the truth of q is taken as an indication of the truth of p, I fail to see that the universal statement is needed. This is, after all, a world in which a very strong positive correlation has been observed between one's engaging in perfect car-owning behavior and his owning the car he drives, which is what matters for the example to be effective.

[29] In this presentation of what matters to my discussion of Moore's paradox in Klein's defeasibility theory, I assume that what qualifies as a blocker need not be what ultimately counts as evidence against p—it need not be the *effective* blocker of the path to p, but may be only what *initiates* the blocking of such a path by initiating a path to the effective blocker. For a discussion of the distinction between *initiating* and *effective* blockers, see Klein 1981.

[30] The reader should expect my adoption of essential elements of Klein's theory to be compatible with an account of *weak* inconsistency according to which one can rationally hold the prefatory belief. Such an account has been provided by Klein himself in his 1985.

Thus, Klein has given us a notion of internal clearance and the Rule of Revision.

Now, suppose that either (p&q)[W]p or (p&q)[W]q when neither p nor q is a contradiction. This is a special case of internal blocking. In such cases, the blocker for either p or q is the other conjunct. This is, for instance, the case of "The sky is clear, but it will rain in five seconds". In our world (or for agents whose belief systems include the meteorological laws we take to be true), this conjunction fails to warrant either of its conjuncts, because they are each other's counterevidence. The conjunction is epistemically self-defeating: it is a case of self-blocking. And a self-blocking proposition is one for which there can be no non-overridden evidence. To see this, consider the case in which the evidence for a conjunctive belief is evidence for each conjunct. Suppose that my evidence for believing "The sky is clear[p], but it will rain in five seconds[q]" is provided by my belief in the proposition expressed by "The weatherman has asserted that p&q" (e). Assuming that assertion distributes over conjunctions, and that e is both evidence for p and evidence for q, my evidence for p&q is its own overrider: since e initiates a warrant path to p, and since (e&p)[W]q, the warrant path from e to q is blocked by e itself (because of the path from e to p)—and the same applies to the warrant path from e to p (because of the path from e to q). So, when evidence for a conjunction is evidence for the conjuncts, but the conjunction is self-blocking, there can be no non-overridden e such that both eWp and eWq. Clearly, however, it would be wrong to assume that it must always be the case that evidence for a conjunction is evidence for the conjuncts—since we want to allow the conjuncts themselves to be evidence for the conjunction. But, whenever the conjuncts of a self-blocking conjunction are the evidence for the conjunction, they can't be non-overridden evidence, since the Rule of Revision precludes their coexistence in one's belief system.[31]

Now, if we are entitled to a distinction between *contingent* and *necessary* self-blocking, I can put forward the following definition of *justifiability*:

A proposition p is *justifiable* if and only if it is possible that, for some S, there is an internally clear warrant path to p. (That is, a proposition p is *justifiable* iff it is not necessarily self-blocking.)

A proposition is *unjustifiable* if and only if it is not justifiable according to the above definition.

Thus, a contingently self-blocking proposition such as 'The sky is clear, but it will rain in five seconds' is justifiable, since it is logically possible that, for some S (one who is significantly different from us), its conjuncts are

---

[31] We have just seen that Klein's 1985 (p. 117) claim according to which, "if 'x' and 'y' range over contingent propositions and if x strictly implies y, then xWy" has to be rejected.

warranted by it. As the example shows, while it must be irrational to believe an unjustifiable proposition, our definition of justifiability allows for the fact that one may also be irrational in believing a justifiable proposition, which is exactly what we want, since *all cases of belief in self-blocking propositions must be cases of irrationally held belief.*

The foregoing suggests that we can speak of "self-blocking beliefs" to describe belief in self-blocking propositions—and there should, of course, be no objection to this use of the term 'self-blocking belief'. But this term should be given a larger extension, as we can see from consideration of the fact that, in some cases (to be considered shortly), even though the proposition itself is not self-blocking, one cannot believe it without acquiring evidence against belief in the proposition. These are cases in which *effective* counterevidence—non-overridable evidence ("unblockable evidence", we might say) that the inclusion of the proposition in the agent's belief system offends against observance of Klein's Rule of Revision—inevitably arises from the very act of belief in the proposition. Let us refer to such a phenomenon with the label 'epistemic self-defeat'. This broader notion of a self-blocking belief—one that accounts for both belief in self-blocking propositions and epistemically self-defeating belief—is captured by the following definition:

> A *belief* b is *self-blocking* for an agent S if and only if either b has a self-blocking proposition as an object or it is impossible for S to include the (propositional) object of b in S's belief system without violating the Rule of Revision.

Since every *contingently* self-blocking proposition is justifiable, and every unjustifiable (necessarily self-blocking) proposition is such that it cannot possibly be included in one's belief system without violation of the Rule of Revision, we have found a class of beliefs—including both beliefs in unjustifiable propositions and epistemically self-defeating beliefs, and no others—that can be defined as follows:

> A *belief* b is *unjustifiable* if and only if it is impossible that, for some S, the object of b be included in S's belief system without violation of the Rule of Revision.[32]

---

[32] Therefore, according to our definitions, not all unjustifiable beliefs have unjustifiable propositions as their objects; but every belief the object of which is an unjustifiable proposition is unjustifiable. It should also be clear that some self-blocking beliefs are justifiable: those the objects of which are *contingently* self-blocking propositions. Thus, some justifiable beliefs are irrationally held; and all unjustifiable beliefs are irrationally held.

It will now be seen that both contradictions and *some* Moore-paradoxical propositions are the objects of unjustifiable beliefs, and that they are the *only* propositions which are the objects of unjustifiable beliefs, which is what makes (some) Moore-paradoxical propositions *contradiction-like* on my account of the matter.

That belief in a contradiction is unjustifiable is easy to see. As noted above, every justified proposition is such that there is a clear warrant path to it. Given this principle, however, contradictions cannot be justified. For reasons that are now familiar to us, no evidence e could possibly warrant both p and ~p. A contradiction is a necessarily self-blocking proposition, and, as we have seen, a self-blocking proposition is one for which there can be no non-overridden evidence. We have also seen that a self-blocking proposition is such that it cannot provide evidential support for all of its logical consequences. Thus, p&~p cannot warrant p, since the warrant path from the contradiction to p is internally blocked by the contradiction itself: the contradiction initiates a warrant path to ~p. Likewise, the contradiction cannot warrant ~p because it initiates a warrant path to p.[33]

It is important to guard against misunderstanding here. Since I have admitted that necessary falsehoods can be rationally believed (and this implies that they can be justified for a given agent), it may seem that this conflicts with my claim that contradictions are unjustifiable (and, so, not justified for any agent). In fact, there is no such conflict. While I am, indeed, committed to the view that *formal contradictions* (propositions which are truth-functionally equivalent to instances of the form p&~p) are unjustifiable—and cannot, therefore, be rationally believed—I need not be committed to the claim that *every* necessary falsehood is unjustifiable, since, according to my definition, unjustifiability arises from violation of the Rule of Revision (either by necessary internal blocking or by epistemic self-defeat), and it is not at all obvious that every necessary falsehood is such that the path to it must be ultimately blocked. In fact, it seems clear that, in many cases, such a path will not be blocked. (Recall the instances of rationally believed necessary falsehoods in section 3 above.) Any residual discomfort here stems, I think, from a tendency to confuse merely having evidence for belief in a proposition with being justified in believing the proposition. There may be the tendency falsely to believe that an agent may rationally believe a contradiction just because he has a reason to believe the contradiction. (Again, a contradiction is a self-blocking proposition, and, as we have seen, there can be no non-overridden evidence for belief in a self-blocking

---

[33] The reason why contradictions can neither be warranted nor provide warrant for other propositions can be found in Klein 1985. He lacks the broader notion of a self-blocking proposition, though.

proposition. So, whenever one has evidence for belief in a contradiction, that evidence is overridden by the contradiction itself.[34])

Like contradictions, all Moore-paradoxical propositions *of a certain type*, though contingent (let us not forget that they can be true!), are such that belief in them is unjustifiable. We are now in a position to understand why this is true of all of (1)–(9), *with one exception* (shortly to be disposed of).

In the case of (1), the reason for thinking that belief in an instance of p&B~p is unjustifiable is provided by the fact that B~p indicates the presence of ~p in my belief system. If I have the higher-order belief that ~p is in my belief system, I surely am in possession of a reason—an overridable reason— to refrain from believing that p. Thus, the right-hand side of the conjunction gives me a reason not to believe the left-hand side. If I believe the conjunction—and, by distribution, believe the conjuncts—I include both p and a reason not to believe p in my belief system, thus violating the Rule of Revision. Notice: the point is not that, in believing a *true* conjunction, I become strongly inconsistent (which I certainly do). Since satisfaction of Condition R does not require that the agent believe truly (even belief in a necessary falsehood may be rational!), the point is that the conjunction itself—*true or false*—gives me a reason (B~p) not to believe one of the conjuncts (p). So, no reason that I might conceivably have for believing the conjunction could possibly override my reason for disbelieving the conjunction, since this is given me with the very act of believing the conjunction. Belief in (1) is epistemically self-defeating.

In the case of (2), to see that belief in any instance of p&~Bp is unjustifiable, we must recall, with Robert Audi, that introspection is traditionally regarded as a source of justified beliefs.[35] Essentially, this amounts to the following: if I believe that p, then, given the availability afforded to us by *possible* introspection (i.e., the fact that introspection is a source of epistemic justification) I have a reason—an overridable (or "blockable") reason—for believing that I believe that p (which, of course, I may never do). Now, if I believe (an instance of) p&~Bp, then, by the distribution principle, I believe that p. But, given that the occurrence of this first-order belief provides me with evidence for the belief that I believe that p, if I believe the conjunction (p&~Bp), I acquire evidence for believing the contradictory of the right-hand side of the conjunction (~Bp). The point here is not that I refute my own belief in believing p&~Bp (which I certainly do, but, again, one may rationally believe a falsehood): the point is that belief in

---

[34] Notice that this is a result of our premises. We haven't just assumed that it must be irrational to believe a contradiction.

[35] I refer the reader to Audi's discussion of this claim about introspection because I think his endorsement of the claim comes with all the right qualifications. See Audi 1998, chapter 3.

the conjunction necessarily furnishes me with a reason to disbelieve the right-hand side of the conjunction. Belief in (2) is epistemically self-defeating.

(3) and (4) seem to pose no special problem: belief in them ought to be considered unjustifiable because they imply a proposition belief in which is unjustifiable, namely (2).

In the case of (6), letting 'Jp' stand for 'I'm justified in believing that p', we see that belief in the conjunction p&~Jp is unjustifiable when we notice that belief in the right-hand side of the conjunction (~Jp) overrides any evidence one might otherwise have for believing the left-hand side (p). Given any piece of confirming evidence e for p in my belief system, conjoining e with my belief that any piece of evidence for p is overridden in my belief system—which is what ~Jp entails—nullifies the evidential support e would otherwise provide for p: (e&~Jp)[W]p.[36] Given our abbreviations, the operative principle here is the following: Jp⊃~B~Jp. Notice, we cannot let evidence be effective when, *by the believer's own lights*, it is not effective! A theory of justification which makes room for epistemic defeasibility—and, according to Thomas Senor, *no* theory can be taken seriously without doing so—*must* accommodate the overriding impact of the believer's own beliefs over whatever else he may believe.[37] We certainly feel that the rational agent is required to be somehow sensitive to the impact of *all* of his (occurrent) beliefs over his ability to include any new beliefs in his doxastic system. (Which is not at all to be involved with any of the familiar controversial claims which my claim may bring to mind: BonJour's "doxastic presumption" or any other claims according to which either higher-order belief or metajustification is a necessary condition of epistemic justification.[38])

Indeed, a denial of the principle that Jp⊃~B~Jp doesn't come cheap. The conceptual viability of epistemological internalism itself seems to depend on such a principle. According to Chisholm (1989, p. 76), the epistemological internalist is one who "assumes that, merely by reflecting upon his own conscious state, he can...find out, with respect to any possible belief he has, whether he is *justified* in having that belief". This (on a JTB+ account of

---

[36] A complication arises here—but one which can be only briefly mentioned, since it does not affect my argument about the form of irrationality exemplified by belief in an instance of p&~Jp. The discussion now shows that my use of '[W]' does not completely coincide with Klein's. Because—like most everyone who's written on the topic—Klein developed his account of how the *overriding* of evidence occurs without considering the form of irrationality we call "Moore's paradox", his use of '[W]' fails to capture the case in which overriding is effected by propositions which would not normally be regarded as *counterevidence* for the propositions whose support in one's doxastic system they destroy. Thus, while I understand that Klein's account of warrant (and his use of 'W') is tenable, I contend that his account of epistemic defeasibility is unduly narrow, in that it conflicts with what is said in my analysis of p&~Jp.

[37] See Senor 1996.

[38] For an account of his "doxastic presumption", see BonJour 1985.

knowledge) implies a commitment to the principle according to which, if I know that p, I'm in a privileged position to know that I know that p. In denying that $Jp \supset \sim B \sim Jp$, one accepts the possibility of his being justified while believing that he isn't. But being in such a state, of course, implies that one may know without being in a position to know that he knows (in a way which has nothing whatsoever to do with whether one possesses the relevant concepts).

It is interesting to notice that the overriding effect of my belief that $\sim Jp$ is *ineradicable*: in believing that $\sim Jp$, I immunize myself against the possibility of a justified belief in p. Therefore, if I believe that $\sim Jp$, this is a self-verifying belief for me. Can *you* truly believe that I falsely believe that I'm not justified in believing that p? Curiously enough, in this case, not even God can second-guess me! If I believe that there is no non-overridden evidence for a given proposition in my belief system, I cannot be wrong: that sweeping condemnation of any possible evidence for the proposition becomes an all-purpose overrider of any potentially sufficient evidence for the proposition.

Finally, notice that there can be no non-overridden evidence for belief in an instance of $p \& \sim Jp$ regardless of whether the believer is justified in believing the right-hand side of the conjunction: even if he is not, there is no non-overridden evidence for belief in the conjunction, simply because, in this case, the evidence for one of the conjuncts (namely, $\sim Jp$) is either non-existent or overridden.[39]

As for (7), on the assumption that $J \sim p$ entails $\sim Jp$, essentially the same argument about the unjustifiability of (6) shows that one cannot justifiably believe an instance of $p \& J \sim p$: the right-hand side of the conjunction overrides any potentially effective evidence one might otherwise have for p in his belief system.[40]

Unlike the Moore-paradoxical propositions previously considered, which were cases of epistemic self-defeat, (6) and (7) are self-blocking. And, since the self-blocking nature of both (6) and (7) depends only on features of the concept of epistemic justification, rather than on any contingent feature of this world, (6) and (7) are necessarily self-blocking.

At this point, the arguments which show that both belief in (8) and belief in (9) are unjustifiable should seem fairly simple. If you believe 'I have no beliefs now', our assumption about introspection ensures that you acquire evidence for believing 'I believe that I have no beliefs now'. This second-

---

[39]   We now see why the objection brought against Olin's conclusion in note 8 above fails: if I am justified in believing that I am not justified in believing that p, then, *really*, I am not justified in believing that p. But she may not have noticed that I am not justified in believing that p even when I am not justified in believing that $\sim Jp$ but believe that $\sim Jp$.

[40]   Unlike $\sim Jp$, (an instance of) $J \sim p$ is not self-verifying. But, in any case, it is not the self-verifying nature of $\sim Jp$ that accounts for its overriding effect.

order belief, however, provides warrant for the first-order belief which is the contradictory of 'I have no beliefs now', namely 'I have some belief now' (logically implied by that second-order belief). Therefore, it is not possible for you to both have non-overridden evidence for believing (8) and believe it. Belief in (8) is epistemically self-defeating.

Now, consider (9). If you assume that existing is a necessary condition of having beliefs, it becomes clear that, if you believe (9), you acquire evidence for the belief that you exist, since, in believing (9), you acquire evidence for the second-order belief that you believe that you don't exist, and this, according to your assumption, entails that you exist. So, as long as you assume that, if you have beliefs, you exist, belief in (9) has to be considered epistemically self-defeating.[41]

None of the foregoing, however, gives us an explanation of the oddity of (5). But we already are in a position to see that all of (1)–(9), except (5), are *absolutely Moore-paradoxical* according to the following definition:

A proposition p is *absolutely Moore-paradoxical* for an agent S if and only if p is not a contradiction and is such that S's belief in p is unjustifiable.

The above definition, in turn, when coupled with Sorensen's notion of an agent's being the *addressee* of a proposition (introduced in section 1 above), lets us explain the absurdity of (5) by including it in the larger class of *Moore-paradoxical propositions*, according to the following definition, which encapsulates my proposed solution to Moore's paradox.

A proposition p is *Moore-paradoxical* for an agent S if and only if either p is absolutely Moore-paradoxical for S or p is such that, necessarily, if p is addressed to S and S believes p, then S believes an absolutely Moore-paradoxical proposition in virtue of believing p.

6. Concluding remarks

In a well-known passage of "On Denoting", Russell claims that "[a] logical theory may be tested by its capacity for dealing with puzzles, and it is a wholesome plan, in thinking about logic, to stock the mind with as many puzzles as possible, since these serve much the same purpose as is served by experiments in physical science". Most of us, I trust, will agree that Russell

---

[41] Peter Klein has shown me that my premises lead to the conclusion that both belief in (8) and belief in (9) are unjustifiable. Clearly, at all events, I am not concerned with putting forward the claim that existing is a necessary condition of having beliefs, even though the truth of such a claim seems obvious enough. My concern is with showing that those who find belief in (9) somehow absurd have at their disposal a compelling argument to the effect that such a belief is unjustifiable.

might as well be speaking of philosophical theories in general, and that the excitement in the philosophical discussion is, for the most part, if not exclusively, derived from those who heed his advice (or, at any rate, behave as if they did).[42] (Even those who think that—or, at least, pay lip service to the view that—philosophers have no business theorizing, or explaining, or testing anything apparently do more theorizing and explaining and testing arguments for soundness than they care to admit.) More to the point, most of us will agree that some puzzles deserve to be classified as "epistemic". These usually include the Lottery, Preface, Knower, and Prediction paradoxes, the paradoxes of analysis, and of confirmation, and Moore's paradox, among others. The jury is still out on how some of these so-called "epistemic puzzles" should best be classified—with philosophers of mind and philosophers of language and philosophical logicians claiming some of these puzzles as their own. Never mind the fact that labeling philosophical problems carries little, if any, explanatory weight. What I want to suggest is that, except for my own proposed solution to Moore's paradox put forward here, the reason one finds in the literature for classifying this puzzle as "epistemic" is as feeble as the claim that the puzzle puts into question the set of necessary conditions of (epistemic) rationality—a claim so generic, so uninformative that it might be used to call the Liar, for instance, "an epistemic puzzle". I submit that my analysis of Moore's paradox gives us a more substantial reason for thinking of the problem as one which belongs to epistemology—and I believe it should be interesting to see what epistemologists can do with the suggestion that the paradox may be intimately connected with some of the easily identifiable problems of the field.

The above quotation from Russell is also intended to motivate the identification of a more specific target for these concluding remarks. I cannot, here, engage in a lengthy speculation about the impact my analysis of the paradox, if correct, may have on the debate between internalist and externalist conceptions of epistemic justification. One potential consequence of my proposal, however, readily comes to mind and can be quickly presented.

Consider Ernest Sosa's (1991) formulation of what he calls "generic reliabilism": "S's belief that p at t is justified iff it is the outcome of a process of belief acquisition or retention which is reliable, or leads to a sufficiently high preponderance of true beliefs over false beliefs". As we know, much of the epistemological activity of the last three decades has been concerned with either searching for the best way to protect the intuitions that motivate generic reliabilism, on the externalist side of the debate, or with learning how to deal with the threats posed by such attempts to improve on

---

[42]   I certainly don't mean to imply that I subscribe to all of Russell's metaphilosophical views. I don't.

generic reliabilism, on the internalist side. In this regard, an observation comes out of my proposed solution to Moore's paradox. It is high time reliabilists who care about Russell's advice at all became concerned with the fact that, if I am right, a conceptual framework that is distinctively internalist may have provided us with the solution to a puzzle which, on most accounts, falls within the epistemological territory, whereas, to my knowledge, we don't have the foggiest idea of what a reliabilist approach to this puzzle might look like. But, it seems to me very clear that a *non-question-begging* reliabilist response to Moore's paradox would probably drive itself into irrelevance by turning its back on the core intuition of generic reliabilism: the idea that the reliability of the belief-forming/sustaining process is *sufficient* for justification. Reconsider a familiar scenario suggested by our proposition (5). My very reliable psychoanalyst tells me: "Your father loves you, but you don't believe he does". With Sorensen, I contend that this is a Moore-paradoxical proposition. Contrary to what the generic reliabilist would have us believe, if I believe what the psychoanalyst tells me, the resulting belief is most definitely *not justified*. No matter what you build into the notion of justification, no matter what you call the epistemic property that is supposed to turn true belief into knowledge, you *cannot know* that your father loves you but you don't believe he does. Generic reliabilism simply gives us the wrong result here. But, if you sever this connection with generic reliabilism, how reliabilistic can you get?[43]

References

Audi, R. 1998. *Epistemology: A Contemporary Introduction to the Theory of Knowledge*. London and New York: Routledge.

Baldwin, T. 1990. *G. E. Moore*. London and New York: Routledge.

BonJour, L. 1985. *The Structure of Empirical Knowledge*. Cambridge, MA, and London: Harvard University Press.

Chisholm, R. 1989. *Theory of Knowledge*, 3rd edn. Englewood Cliffs, NJ: Prentice Hall.

Collins, A. W. 1996. "Moore's Paradox and Epistemic Risk". *The Philosophical Quarterly* 46, pp. 308–19.

de Almeida, C. 1998. "Klein on the Defeasibility Theory", *Crítica* (Universidade Estadual de Londrina, Brazil), vol. 3, no. 11, 225–37.

---

[43] My greatest debt is to Peter Klein—*without whom not,* as they say. As I see it, all those who have written on Moore's paradox in the last twenty years or so owe Roy Sorensen and John N. Williams a debt of gratitude for their groundbreaking work on the problem. I am personally indebted to Roy for stimulating discussion of it. I have also benefited from comments by Chad Mohler, Paulo Faria, Cesar Mortari, and two anonymous referees for this journal. An early version of the paper was presented at the 4th Analytic Philosophy Conference, in Florianopolis, Brazil. I am grateful to my audience on that occasion, particularly to Oswaldo Chateaubriand and David B. Martens. This project was partially funded by CAPES.

_____. 1999. "Moore's Paradox". In R. Audi, ed., *The Cambridge Dictionary of Philosophy*, Second Edition (Cambridge: Cambridge University Press).

Etchemendy, J. 1995. "Semantic Paradoxes". In R. Audi, ed., *The Cambridge Dictionary of Philosophy*. Cambridge: Cambridge University Press.

Foley, R. 1993. *Working Without a Net: A Study of Egocentric Epistemology*. Oxford and New York: Oxford University Press.

_____. 1995. "Pragmatic Contradiction". In R. Audi, ed., op. cit.

Goldstein, L. 1988. "Wittgenstein's Late Views on Belief, Paradox and Contradiction". *Philosophical Investigations* 11:1, pp. 49–73.

Gombay, A. 1988. "Some Paradoxes of Counterprivacy". *Philosophy* 63, pp. 191–210.

Grice, H. P. 1961. "The Causal Theory of Perception". Partially reprinted in his *Studies in the Way of Words*. Cambridge, MA: Harvard University Press, 1989.

Heal, J. 1994. "Moore's Paradox: A Wittgensteinian Approach". *Mind* 103, pp. 5–24.

Klein, P. D. 1981. *Certainty: A Refutation of Scepticism*. Minneapolis: University of Minnesota Press.

_____. 1985. "The Virtues of Inconsistency". *Monist* 68, pp. 105–35.

_____. 1986. "Immune Belief Systems". *Philosophical Topics* 14, pp. 259–80.

_____. 1996. "Warrant, Proper Function, Reliabilism, and Defeasibility". In J. L. Kvanvig, ed., *Warrant in Contemporary Epistemology: Essays in Honor of Plantinga's Theory of Knowledge*. New York and London: Rowman & Littlefield.

Lehrer, K. 1990. *Theory of Knowledge*. Boulder and London: Westview Press and Routledge.

Linville, K. and Ring, M. 1991. "Moore's Paradox Revisited". *Synthese* 87, pp. 295–309.

Malcolm, N. 1995. "Disentangling Moore's Paradox". In his *Wittgensteinian Themes: Essays 1978–1989*, ed. by G. H. von Wright. Ithaca and London: Cornell University Press.

Moore, G. E. 1942. "Reply to My Critics". In P. Schilpp, ed., *The Philosophy of G. E. Moore*. La Salle, Ill.: Open Court.

_____. 1944. "Russell's 'Theory of Descriptions'". In P. Schilpp, ed. *The Philosophy of Bertrand Russell*. La Salle, Ill.: Open Court.

_____. 1993. "Moore's Paradox". In *G.E. Moore: Selected Writings*, ed. by T. Baldwin. London and New York: Routledge.

Olin, D. 1983. "The Prediction Paradox Resolved". *Philosophical Studies* 44.2, 225–33.

Plantinga, A. 1993. *Warrant and Proper Function*. Oxford and New York: Oxford University Press.

_____. 1996. "Respondeo". In J. L. Kvanvig, ed., op. cit.

Senor, T. 1996. "The Prima/Ultima Facie Justification Distinction in Epistemology". *Philosophy and Phenomenological Research* 61, No. 3, 551–66.

Shoemaker, S. 1988. "On Knowing One's Own Mind". In James E. Tomberlin, ed., *Philosophical Perspectives 2: Epistemology*. Atascadero, CA: Ridgeview.

___. 1996. "Moore's Paradox and Self-knowledge". In S. Shoemaker, *The First-Person Perspective and Other Essays*. Cambridge: Cambridge University Press, pp. 74–93.

Sorensen, R. A. 1988. *Blindspots*. Oxford and New York: Oxford University Press.

___.1992. "Moore's Paradox". In J. Dancy and E. Sosa, eds., *A Companion to Epistemology*. Oxford: Basil Blackwell.

Sosa, E. 1991. "Reliabilism and Intellectual Virtue". In his *Knowledge in Perspective: Selected Essays in Epistemology*. Cambridge: Cambridge University Press.

Williams, J. N. 1994. "Moorean Absurdity and the Intentional 'Structure' of Assertion". *Analysis* 54, pp. 160–66.

Wittgenstein, L. 1953: *Philosophical Investigations*. Oxford: Basil Blackwell.

___. 1980. *Remarks on the Philosophy of Psychology*, vol. I. Oxford: Basil Blackwell.