



Deposited via The University of Leeds.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/106799/>

Version: Accepted Version

Book Section:

Woods, JE (2018) A Commitment-Theoretic Account of Moore's Paradox. In: Turner, KP and Horn, L, (eds.) Pragmatics, Truth and Underspecification; Towards an Atlas of Meaning. Current Research in the Semantics / Pragmatics Interface (CRISPI), 34. Brill, pp. 323-348. ISBN: 978-90-04-36544-5.

<https://doi.org/10.1163/9789004365445>

© 2018, by Koninklijke Brill NV, Leiden, The Netherlands. This is an author produced version of a book chapter published in Pragmatics, Truth and Underspecification: Towards an Atlas of Meaning. Uploaded in accordance with the publisher's self-archiving policy.

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

A Commitment-Theoretic Account of Moore's Paradox*

Penultimate Version of a paper forthcoming in a volume on semantics and pragmatics. Please cite the published version when it appears.

Jack Woods

1. Introduction

Moore's paradox, the infamous felt bizarreness of sincerely uttering "I believe grass is green, but it ain't" and things of similar form, has attracted attention since its original discovery (Moore 1942).¹ Moore-paradoxical utterances *feel* contradictory, even though both conjuncts can obviously be true. It is often taken to be a paradox of *belief*—in the sense that the locus of the inconsistency is the beliefs of someone who so sincerely utters. This claim has been labeled as the *priority thesis*:

If you have an explanation of why a putative content could not be coherently believed, you thereby have an explanation of why it cannot be coherently asserted. (Shoemaker 1995, 227)²

The priority thesis, however, is insufficient to give a general explanation of Moore-paradoxical phenomena (it's also false) as:

I promise I'll marry you, but I won't

*Thanks to Sarah Raskoff for the discussion which prompted this piece, Paul Benacerraf and John Burgess for useful discussions on the topic of Moorean absurdity, and Derek Baker, Eliot Michaelson, and Ken Turner for very useful comments.

¹Philosophers have often distinguished the omissive form of Moore's paradox— p , but I don't believe that p —and the commissive form— p , but I believe that not p . I deal primarily with the omissive form here, but I don't think there is much philosophical traction to be gained out of the distinction once we reject the idea that we explain Moore's paradox by means of inconsistent or irrational belief. The commissive form more or less transforms into the omissive by means of the presumption (not always true!) that quite generally people believe not p only when they don't believe p .

²Atlas (2007) has argued vigorously (and, I think, successfully) against Shoemaker's account on grounds that Shoemaker illicitly moves from "manifesting a belief in P " to "manifestly asserting my belief in P ". He also argues against the uniform paradoxicality of constructions of the form " F , but I don't believe that F " by constructing cases in which they can be sincerely asserted. I agree with the examples, but think that they can be handled by the account I suggest at the close of this paper.

has the same paradoxical character as:

I believe it's raining, but it ain't

even though the latter, but not the former, involves expression of the utterer *believing* thus and so. Moore-paradoxical utterances can occur in non-assertoric constructions and, moreover, can occur in the context of speech acts which do not express mental states at all.³ I will leave aside, for the purpose of this essay, whether we should be giving an independent analysis of Moore's paradox for thought or, alternatively and as I prefer, whether we should regard Moore-paradoxical thought as derivative from Moore-paradoxical speech.⁴ I will not address the priority thesis again directly, but it should be clear to the reader that the considerations below tell directly against it.⁵

The fact that there are Moore-paradoxical constructions which do not involve expressing mental states supports the hypothesis that Moore paradoxicality arises due to speech acts engendering *commitments* of a particular sort. Some speech acts, like typical assertions, give rise to commitments to *believing* thus and so.⁶ Others, such as promises, give rise to *acting* thus and so. When we go on to immediately deny that these commitments are met, our purpose in engaging in the speech act becomes mysterious and interpreters quite reasonably become puzzled about what we were attempting to accomplish by means of our utterances. We have seemingly undermined the purpose of making the promise or assertion at all. Basic linguistic competence, presumed of those we typically interpret, requires mastery of the commitments incurred by asserting, commanding, and promising, and the like.⁷ Given this, explicit claimed violation of these commitments gives rise to our inability to successfully interpret the speaker. This is the basic sense of "inconsistency" or "oddness" that we now call Moore paradoxicality.⁸ Or, anyways, so I will argue.

I originally argued for a version of this claim in a paper rejecting expressivism (Woods 2014). It struck me then, as it strikes me now, as the right sort of take on a general phenomenon which had been typically addressed in only its belief-involving manifestations. I was happy to find afterwards that (Atlas 2007)

³See (Williams 2013) for independent arguments against the priority thesis. Williams focuses on the failure of the priority thesis when we restrict our attention to the case of belief. My interest is in showing that even if the priority thesis holds for belief, it doesn't suffice to explain other cases of Moore paradoxicality.

⁴See (Baker and Woods 2015, §IV.B) for suggestions in this direction.

⁵For good examples of linguistically acceptable utterances plausibly manifesting irrationality, like "I shouldn't do it, but I will anyway", see the discussion in (Baker 2014, §6).

⁶Take 'typical' here to be meant seriously; I do not mean to deny that there are contexts where a more relaxed standard for asserting is in force. After all, we've all shot the shit a time or two.

⁷This, of course, is a familiar view. For *loci classici*, see (Searle 1969) and (Brandom 1994). For a more recent discussion treating of related issues that arise with lying, see (Michaelson 2016) and (Stainton 2016).

⁸I mean 'inconsistency' here in the older and broader sense of "standing in tension with one another", not the narrower "unable to both be true".

provided further and distinct reasons useful to establishing what I originally took the most natural explanation of Moore paradoxicality. Atlas writes:

... it is bizarre to think that the impossibility of the rational, sincere assertion of a Moore sentence could explain the oddity of uttering a Moore sentence assertively (as Moore (1993: 207) would put it), even if the impossibility of a sincere, “rational” assertion explains why either the utterance is insincere, or expresses a “non-rational” belief or a corrigible belief, or is not an assertion—oddities all, but none of insincerity, non-rationality, corrigibility, and an assertoric speech-act “misfire” is the oddity of uttering the Moore sentence assertively. None of those consequences would seem to explain the particular oddity of uttering the Moore sentence assertively. (As Paul Benacerraf once remarked, in reaction to this claim, an adequate account of the oddity must appeal to principles that every speaker/hearer can be expected to “know”, in the sense in which we know our language and its uses. Not every speaker, surely, is natively a Port Royal logician.) (Atlas 2007: 142-143)

and I completely agree. These sorts of explanation do not get at the root of the problem. Insincere assertions—“I’m pleased to see you, but I ain’t”, when I ain’t—still give rise to Moore paradoxicality, as do irrational and corrigible assertions. The madman who yells that the sky is constantly falling in, but then immediately demurs is not just mad, but linguistically incompetent. And failures to assert at all tend to undermine the phenomena in question since there is a natural and immediate explanation of why the failed assertors have failed to “contradict” themselves in the relevant sense—they have not actually made an assertion. What we need is a way to chisel down to the class of utterances where Moore paradoxicality arises and, then, to give an actual explanation of why, in many cases, these utterances give rise to such a feeling of oddness.

Since insincere promises, commands, and so on can still commit, commitment violation by itself is insufficient to explain Moore paradoxicality. What we need is a *general* account of *why* we get Moore paradoxicality in some cases—one which explains why these cases are puzzling for everyone, linguistically savvy or not. We also need to explain why it is so easy to multiply Moore-paradoxical cases for a variety of different speech acts. We can articulate these demands in terms of two desiderata of an explanation of Moore paradoxicality:

UNIFORMITY: An explanation of Moore paradoxicality must explain its occurrence in all cases, not merely the somewhat special, if historically important, case of first-personal assertions.

EGALITARIANISM: An explanation of Moore paradoxicality must explain its manifestation to all linguistically competent interpreters of speech (when interpreting speakers themselves presumed linguistically competent.)

Since the irrational and the insincere typically count as linguistically competent, it isn't sufficient explanation of a Moore paradoxical utterance that the utterer is seen as irrational or insincere *as a result* of the seeming inconsistency or irrationality of their utterance. In what follows, I'll first argue for the breadth of Moore-paradoxical phenomena. I will then argue that this breadth makes implausible Shoemaker's attempt to derive Moore paradoxicality from the "logic" of belief. Similar views will fall for similar reasons, though I do not argue for this explicitly. I will then turn to developing the commitment-theoretic explanation of Moore's paradox I favor.

The key to my explanation involves welding a general theory of the commitments arising from speech acts to a constraint on normal interpretation of the utterances of others. The constraint, roughly, is that utterers should generally be taken to reflexively intend to take on the commitments incurred by means of the speech act they are engaged in. In particular, this means that when a speaker says "grass is green, but I don't believe it is", they reflexively intend to take on the commitments of asserting that grass is green. Since these commitments involve believing what I assert, when I deny this commitment in the second conjunct, I undercut my presumed intention to take on this commitment and thereby puzzled the speaker. I seem, in the most literal sense, inconsistent. Likewise, when I say "I promise to give you 5 Turkish lira, but I won't", I am presumed to reflexively intend to take on the commitments of promising which include giving you 5 Turkish lira. Only presumed, of course, since I might actually be insincere. So, when I claim that I won't fulfill this commitment, I give rise in my interpreter the puzzling feeling of directly undermining my earlier promise.

This account allows me to meet the constraint that flagrantly insincere assertions, promises, and what have you can be felicitously uttered—as they obviously can be. Irrationality is no excuse for failing to meet a promise, in many cases, and it certainly is no excuse for not believing what you assert. What is needed for cases with the form of Moore's paradox to be linguistically felicitous is that the default presumption of reflexive intention be waived in that context. I will discuss such cases below (see §2.2, §5). I conclude that the resulting account satisfies UNIFORMITY, EGALITARIANISM, and doesn't over-predict occurrences of Moore's paradox. What more could we want from an explanation?

2. Much Moore Paradoxicality

Sincerely uttering any of the following in non-deviant contexts gives rise to the feeling we call Moore paradoxicality:

- (1a) It's raining, but I don't believe it is.
- (1b) Get me a beer, but I don't want you to get me a beer.
- (1c) Boo Yankees! I have no negative feelings towards the Yankees.

- (1d) Would that I could have another drink, but I don't wish for another.
- (1e) I promise I'll pick you up at the airport, but I won't.
- (1f) I swear to you that he's a spy, but he isn't.
- (1g) It might be raining, but it's not.

None of these except the first involve belief in a direct fashion. (1b) involves denying the possession of a desire, (1c) involves denying possessing certain evaluative attitudes, (1d) involves denying that I wish in certain ways, (1e) involves denying that I'll act in certain ways, (1f) involves denying that a certain fact holds, and (1g) the same.

Note, as evidence of the Moore paradoxicality of these constructions, that switching to the past tense (where it makes sense) results in sensible constructions:

- (2a) It was raining, but I didn't believe it was.
- (2d) Would that I could have had another drink, but I didn't wish for another.
- (2e) I promised I'd pick you up at the airport, but I won't.
- (2f) I swore to you that he's a spy, but he wasn't.
- (2g) It might have been raining, but it wasn't.⁹

The coherence and lack of paradoxical character of the past-tense versions of present-tense oddness is one of the characteristic marks of Moore paradoxicality. The examples above are also non-paradoxical when we put them into the second or third-person (where such a modification is natural).

Are these examples really Moore paradoxical? Yes, it seems so. Searle and Vanderveken (1985) argue that any speech act conjoined with the denial of its sincerity condition will, in typical circumstances, produce the usual reaction we label as Moore paradoxicality. Consider, for example:

What time is it? But I don't want to know what time it is. (Williams 2015)

Searle and Vanderveken do not discuss cases like (1f), but they feel entirely analogous to me and, I hope, to you. Sincerity conditions are a usual and flagrant case in which we get Moore paradoxical phenomena, but it is entirely unclear why they should be the only such cases.¹⁰ The above, *especially* the promissory

⁹(1b) and (1c) do not have natural past tense constructions.

¹⁰For an entertaining non-standard case, consider Sorensen's 'God knows that we are not theists' (Sorensen 1988: 17). On the presumption that sincere utterance of this commits me to the existence of God, this looks like a similar case. On the other hand, it threatens to collapse a potentially useful distinction between pragmatic paradox and Moore's. So I shall not further pursue this type of case here.

cases, seem clearly instances of the phenomenon that Moore described.¹¹

If these examples really are Moore paradoxical—and I put it to you that they are—then we need an account of the phenomenon they are all instances of.¹² That is, satisfying UNIFORMITY demands that our account of Moore paradoxicality encompass assertions, imperatives, ejaculations, optatives, promissory utterances, avowals, and epistemic modals. Moreover, such an account should predict Moore paradoxicality in cases where it occurs, but should *not* predict Moore paradoxicality in cases where it does not or, importantly, assimilate ordinary cases of having inconsistent beliefs to the same phenomena.

EXTENSIONAL ADEQUACY: An explanation of Moore paradoxicality should not *overspill* to cases of ordinary inconsistent or irrational beliefs.

Can we make sense of UNIFORMITY if we explain typical cases of Moore-paradoxical utterances on the basis of inconsistent beliefs? In order to do so, we would have to derive inconsistent beliefs from beliefs expressed by the above constructions. Consider, for example, (1f). Why is it that we get Moore paradoxicality from swearing to a fact, then immediately denying that the fact holds. The potential explanation, on the inconsistent belief approach, holds that when we swear that p , we express a belief that p and when we deny p , we express our belief that *not* p . Presuming that believing that *not* p rationally entails—requires, on pain of irrationality—not believing p , we get that we do not believe p (at least if we are rational). But believing p and not believing p are inconsistent, so we have displayed that we believe in a way which can't be satisfied. And this, presumably, accounts for the feeling of contradiction. This explanation also vindicates the first-personal present-tense character of Moore's paradox. I can coherently say of you that you swore up and down that grass is purple, but it ain't. I can even coherently say of me that I swore up and down that there was no meat in the dish, but there was.

There is a certain unloveliness to this type of explanation. It is not at all clear that in swearing to something or promising to do something we express the sort of beliefs necessary to generate Moore paradoxicality. Even putting this problem to the side, we also need very contentious principles of rational belief in order to carry out this type of explanation.

¹¹I shall not here argue further that these cases are really Moore paradoxical. My account works even if they're not—though its motivations become slightly less clear. (Heal 1977) is often cited as claiming that (1b) and like constructions are not Moore paradoxical, but I read her as less concerned with Moore paradoxicality in general and more concerned with whether *wanting* is a sincerity condition on commanding. It's worth noting that my solution below (§4) neatly handles her examples as all of them are cases where the presumption of a reflexive intention on the part of the utterer is not in force.

¹²I am willing to hedge here and say that for those of us, like myself, Searle and Vanderveken, etc. who find Moore's paradox in cases outside of first-personal assertions, another account is needed. Of course, I think the data I canvassed above is sufficient to establish that Moore's paradox is a more general phenomenon, but that is a fight I cannot have here.

2.1 Problem 1: Contentious Principles of Belief

Consider, for example, that not only (1f), but also:

(1f*) I swear to you that he's a spy, but I don't believe it.

displays Moore paradoxical character. Our explanation has to somehow explain both. If we explain (1f) by appeal to the fact that swearing that p expresses my belief that p , then presumably we should claim that denying that we believe that p expresses our belief that we do not believe p . Presuming that believing that we do not believe that p rationally entails not believing p , we likewise obtain an inconsistency in our beliefs.

We have now invoked a fair number of contentious principles of belief in running our explanation. We need principles of rational belief like:

- $B(\neg p) \Rightarrow_r \neg B(p)$
- $B(\neg B(p)) \Rightarrow_r \neg B(p)$
- $\Rightarrow_r \neg(B(p) \wedge \neg B(p))$

in order to generate the relevant contradiction. But these are not obvious on their face: suppose, for instance, I believed q and thereby came to be required to believe *not* p on pain of irrationality. Does it follow, on pain of irrationality, that I need to refrain from believing p ? Not obviously, especially if I am unaware of the fact that I'm rationally committed to believing *not* p .

These principles are more plausible when taken to govern *obvious rational requirements* for explicit belief, but even there they are not obvious. Is it obvious that if I'm required to believe that I believe p that I am thereby obviously required to believe p ? No; suppose I'm trying to recover my beliefs after a terrible accident. You, a reliable guide to my previous beliefs, testify that I did not believe p , but you chuckle as you say it. I ask why and you mention that even though I didn't believe p , I believed that I believed p and that I defended both stances vehemently. Presumably, if I think I had good reason and I trust you, it's plausible that I am rationally required (given the evidence) to believe that I believe p and, additionally, to not believe p —though here only on the presumed existence of evidence for this stance.¹³ It would be far better to have an account of the problem which does not presume contentious principles of rational belief like the above.

2.2 Problem 2: Irrationality and Moore's Paradox

A more pressing worry, which I articulated in my earlier work, is that irrationality isn't sufficient to offset Moore-paradoxical reactions (Woods 2014, objection

¹³This case was suggested to me by thinking over the example of amnesiac Gray Davis in (Atlas 2007: 131)

3). The utterer of a Moore-paradoxical utterance displays linguistic incompetence, not outright irrationality. We should distinguish these reactions accordingly. Compare, for example, the person who asserts one of:

- I want a cookie, they're in the jar, but there's no reason to go to the jar
- István is more likely a libertarian librarian than a librarian
- Either Saniye or Sam went to the bank, Sam went only if Saniye did, but Sam didn't go

They are explicitly displaying one or another form of irrationality, but it is hard to accuse them of linguistic incompetence. Rather, on the presumption that they mean the same by their words that we would, they fail to be fully rational.

Or, alternatively, consider a direct case. Suppose I'm a known lunatic, likely to spurt this or that nonsense at any turn. Does that make my utterance of:

I'm made of sugar

of my utterance of

Fictional characters can't exist, yet here I am¹⁴

incoherent? I submit the answer is no. They are just resoundingly and obviously false. On the other hand, even if I am a known incoherent, my utterance of:

I believe that I myself am made of sugar, but I'm not

is resoundingly bizarre and feels self-contradictory.¹⁵ Why is this? Clearly, it has nothing to do with displaying the fact that I'm incoherent or irrational. Rather, it has to do with some feature of these cases over and above the display of irrationality.

It should not be surprising that we can find cases of exhibited irrationality that are not Moore paradoxical. Explicit contradictions can be felicitously uttered, but presuming the connection of belief and assertion, such an utterance will commit the utterer to possessing inconsistent and thereby irrational beliefs.¹⁶ Atlas is sensitive to this point:

A difficulty with Langford's explanation is that there are a denumerable number of self-contradictory statements that are perfectly felicitous, so that the logical inconsistency of the total signification of a statement cannot be sufficient for its assertoric oddness. (Atlas 2007: 142)

¹⁴This example is plausibly an example of a *structural* irrationality, unlike the preceding, which is plausibly an example of *substantive* irrationality.

¹⁵I here adopt Atlas's trick of putting in an explicit reflexive pronoun so as to force the right reading. See (Atlas 2007: 136-137).

¹⁶I don't think all inconsistent beliefs are irrational, but even some manifestly irrational inconsistencies can be felicitously uttered.

and we should be as well. Moore’s paradox is interesting and worth investigating *exactly because* it is not simply a case of some fancy type of inconsistency or irrationality in our beliefs or the content of what we assert.

It is also possible to cook up potential cases in the opposite direction; it is not obviously irrational, though it is obviously Moore paradoxical, to assert that the theory of evolution is correct, but that I don’t believe it (at least in the *recherché* sense of irrationality involving asserting against the total balance of evidence.) After all, I might have evidence for believing that the theory of evolution is false, but be unwilling to bring this evidence to bear on what I am willing to assert. If these considerations are right, and if irrationality simply doesn’t imply linguistic incompetence, then we need an account of Moore paradoxicality that doesn’t reduce it to exhibition of irrationality. That is, we need an account that respects EGALITARIANISM.

2.3 The Alternative

Even if we could finesse the above problems with rationality and principles of rational belief, we would face another. UNIFORMITY already makes plausible the idea that we should look elsewhere than inconsistent beliefs for an explanation of Moore paradoxicality. To see why, consider (1e). Why is it Moore paradoxical to promise while denying that I’ll meet that promise? One explanation claims that when we promise, we express a belief that we will fulfill that promise and when we deny we’ll meet that promise, we express a belief that we won’t. However, even if promising to give you money, then denying I believe that I will gives rise to Moore paradoxicality, the case seems very different than promising to give you money, then explicitly denying that I will. And, since promising to give you money, then denying I *intend* to fulfill that promise likewise seems Moore paradoxical, we need to invoke the contentious claim that intentions (or lack thereof) require us to believe that we will (or won’t) satisfy them. That is, if we tried to explain the Moore paradoxicality of the last case on the inconsistency of belief model.

Similar pressure can be found in reconsidering cases of manifest displays of irrationality which nevertheless do not manifest the sense of linguistic incompetence characteristic of Moore’s paradox. Moran (2001: 84) suggests that we find such cases on the therapist’s couch, where through a combination of external evidence and behavioral observation, a subject realizes that they really do believe such and so, but also affirm that they do not, reporting from “the inside”, so to speak. “My mother didn’t mistreat me, but I believe she did”, in such a context, is a display of a manifest psychological irrationality, but one which does not trigger the sort of feeling of linguistic incompetence we see in the above examples. So there seems to be a mismatch between manifesting psychological irrationality and the jarring phenomenon that Moore described.¹⁷

¹⁷Thanks to Derek Baker for helpful discussion. Moran nevertheless suggests that there is

The cases here pile up. We can make some kind of coherent sense of booing a sports team, then later denying that I really believe that they're terrible, whereas booing a sports team and then denying I dislike or disfavor them seems Moore paradoxical. Consider also that I can explicitly exhibit that I fail to meet a commitment by my actions without any direct commitment to believing that I'm failing to meet it. Suppose, for example, I promise to return your favorite shirt to you while simultaneously intentionally and explicitly setting it on fire. I haven't asserted that I believed that I won't satisfy my promise, though I have definitely displayed my lack of intention to do so. Even if we can manage to stuff these constructions under the general belief-oriented account, we do so by giving *indirect* explanations of their Moore-paradoxical character that neglect the character of the particular speech act involved. A better account which respected this difference is preferable.

An alternative view is available: we need merely take speech acts, like swearing and promising, to generate commitments in some way and view overt declaration that we will or do violate these commitments as underlying the linguistic oddness of Moore's paradox. Take our example from above of swearing to the truth of p . This speech act generates a commitment to the truth of p . We could call such a commitment a *guarantee*. When we deny that we meet it, we cause confusion in the interpreter and thereby the feeling of contradiction. There is no contradiction, of course, just inconsistency between what we have committed to and what we have said is the case. How, that is, could I guarantee the truth of something I am willing to assert is false? This style of explanation promises to vindicate both UNIFORMITY and EGALITARIANISM. We might worry, though, that it generates too much Moore paradoxicality and violates EXTENSIONAL ADEQUACY.

2.4 Too Much Moore Paradoxicality?

Adopting some version of this account, we will find Moore paradoxicality all over the place.¹⁸ Consider, for example, the speech act of promising. A typical

a problem with the pragmatic picture of Moore's paradox since "...the prevalent diagnosis of Moore's Paradox as the paradigm of a "pragmatic paradox" is inadequate at best, for the puzzling quality persists outside the context of utterance and speech act." (70). However, his examples of this persistence involve internally tokening the Moore paradoxical thought. This, of course, is an action that may very well bring along commitments all of its own, whether basic or derivative from commitments involved in ordinary utterance or assertion of the corresponding sentence (Baker and Woods, 2015, §4 explores a similar strategy to distinguish Moore-paradoxical incoherence from semantic incoherence). Since I have put aside Moore's paradox for thought above (and since the pragmatic diagnosis is hardly prevalent), I won't pursue this point further here.

¹⁸Although, as I argued in (Woods 2014), we *do not* find it where the expressivist predicts we should. See (Toppinen 2014, 2015) and (Woods 2015) for discussion and (Raskoff ms) for further worries. Baker (2014, 2015) argues that we also do not find it where guise-of-the-good theorists predict we should.

promise plausibly commits me to (a) keeping the promise, (b) intending to keep the promise, and (c) believing that I can keep the promise. Correspondingly, we should expect that:

(3e) I promise to pick you up, but I won't.

(4e) I promise to pick you up, but I don't intend to pick you up.

(5e) I promise to pick you up, but I don't think I can.

feel Moore paradoxical. Which, I submit, they do. Of course, they all feel slightly different in terms of strength and the conditions under which we can felicitously utter each differ, but this is as we should expect. Commitments come in different strengths. This is, in fact, a virtue of the account; it explains in a natural way why:

(1a) It's raining, but I don't believe it is.

feels more Moore paradoxical than:

(6a) It's raining, but I don't have reason to believe that.

and:

(7a) It's raining, but the evidence (on balance) suggests otherwise.

is not Moore paradoxical at all. We are plausibly committed to having *some* reason for asserting what we believe, we are strongly committed to believing what we assert, but we are not at all committed to asserting in line with the balance of evidence. Put a slightly different way, sincerely asserting either of (1a) or (6a) is evidence of linguistic incompetence, whereas sincerely asserting (7a) is merely evidence of epistemic incompetence. The explanation for this is that in (7a) we do not overtly deny in the right conjunct commitments incurred by the left conjunct. Whereas, in (1a) and (6a), we do.

Of course, there may be more cases of overspill. Consider a case in which I promise to do something that everyone knows that I cannot. Or, alternatively, a case in which I promise to do something, the promising of which entails (in light of other background commitments), that I cannot keep it. Are these cases of Moore's paradox? Presumably not; a reasonable account of Moore's paradox should explain this fact. Below, after I discuss in significantly more detail the notion of commitment, I will turn to such cases and show that the addition of a constraint of normal interpretation—that the speaker reflexively intends their utterance to be taken in a certain way—rules out these sorts of cases and allows a more plausible commitment-theoretic account of Moore's paradox.

3. Commitments

The explanation I want to offer makes crucial use of the notion of *commitment*.¹⁹ What are these commitments though? The rough outline of the story I want to tell claims that commitments are the product of a system of *formal* norms governing our speech behavior. Commitments are not to be reduced to the mental state of *feeling committed* or some form of *intention*; rather, they are social facts about what we ought to do given what we have done. I take on a commitment by performing certain actions—those actions the system of norms marks as conditions for being committed. For example, when I promise, I commit myself to satisfying that promise, whether or not I intend to do so or feel like I must. Likewise, when I swear to the truth of something, I am criticizable if it is false, whether or not I believed that it was true when I so swore: I take on a commitment to that thing being true. We may perform these actions knowing full well that in doing so, I take on a commitment. We may perform these actions without knowing any such thing. Sometimes ignorance of the law is no defense, while sometimes ignorance of the law is a full defense; likewise with social facts.

Where, though, do these commitments come from? Drawing on my recent work (Woods forthcoming, manuscript), I take commitments to be the product of implicit social rules, much like I take legal, moral, and promissory norms to be the product of implicit social rules. These social rules give rise to notions of *blame* and *criticizability* which underwrite, in many cases, normative judgments of people who violate them. They also give rise to constitutive rules for analyzing and interpreting the speech of others, analogous to how the social rules governing a game of chess or bar pool give rise to constitutive rules for interpreting the outcome of a game of chess or bar pool. Importantly, in voluntarily taking on a commitment, I *bind* myself to attempting to satisfy it, at least typically, and can be fairly interpreted as taking that commitment as a constraint on my action. Of course, I can be committed in ways which don't guide my action and not all commitments are explicitly recognized, but the (interpreted as) voluntary undertaking of a commitment will be important here and below. When I am under a commitment, it will typically be explained by both the system of social rules giving rise to norms and the performance of the action which put me under the commitment.²⁰

¹⁹This focus on commitments isn't entirely novel to me. Commitments have been used for analyzing illocutionary acts by Searle (1969), Brandom (1983), Wright (1992), Alston (2000), MacFarlane (2005), Harnish (2005), and others. The focus on using them to explain Moore-paradoxical behavior is relatively novel, being invoked explicitly only by Baldwin (2007) to my knowledge. Baldwin, though, focuses on our undertaking a commitment—note the psychological character here—to the truth of what we judge. My analysis is less mentally-oriented and broader, as I countenance commitments to lots of things besides the truth of what we judge or assert.

²⁰Some commitments, of course, are standing and don't require any action other than participation in a particular social context in order to be in force. See (Woods forthcoming, manuscript) for more detail, especially of how to get *reasons to act* out of commitments.

We will start by looking at one of the more common ways of proceeding here, the notion of a sincerity condition, and show that it is inapt as a full analysis of the relevant notion of commitment. I will then turn to a more general picture on which sincerity conditions form a special subclass of the commitments of particular speech acts. It turns out that even commitments are insufficient to generate Moore paradoxicality since we can incur commitments by means of speech acts where we are explicitly incapable of meeting the commitment. I then turn to adapting a proposal of Bach and Harnish (1979) to assuage this problem.

3.1 Sincerity Conditions

Consider the paradigm case of a speech-act engendered commitment—the connection of assertion with belief. When we make an assertion, absent special conditions, those who hear us presume us *sincere*. This means, in particular, that they take us to believe what we assert. If I say “this paper is late”, then unless I’ve given some signal that I’m not to be taken seriously—an upraised eyebrow, a bizarre intonation contour, etc—I am presumed to believe that this paper is late. If I do not believe this, and this comes out, then I am *criticizable* for not speaking sincerely.²¹ If I don’t believe this paper is late, and it doesn’t come out that I don’t, I am nevertheless fairly interpreted as believing this paper is late.

This connection between belief and assertion is often called a *sincerity condition*. That is, we take it that what it is to sincerely assert that my paper is late constitutively involves me believing that my paper is late. Sincerity conditions apply to speech acts other than assertions. Plausibly, it is insincere for me to promise that I will give you 5 Turkish lira if I have no *intention* of doing so and it is likewise insincere for me to request that you get me a beer if I have no *desire* that you do so. Many theorists have taken the expression of a sincerity condition to be fundamental to the explanation of Moore-paradoxical phenomena.

...when one performs the speech act one necessarily expresses the sincerity condition and thus to conjoin the performance of the speech act with the denial of the sincerity condition would be to express and deny the presence of one and the same psychological state. (Searle and Vanderveken 1985: 19)

However, contra these theorists, this does not mean that we fail to commit ourselves in some way when we explicitly insincerely promise or request. Ac-

²¹Strictly speaking, I am *criticism-liable*, not yet criticism-worthy. It may be that criticism is misplaced, depending on the conditions under which I failed to believe what I asserted. Nevertheless, someone is licensed to criticize me (even if it would be a moral error for them to do so) merely by the fact that I have violated a norm of correct assertion. Compare the legal notions of being prosecution liable and prosecution worthy.

ording to many promissory conventions, if I promise you to pay 5 Turkish lira, as I pull out my empty pockets, I have still promised. And, likewise, if I request you get me a beer, while telling my friend that I really don't want you to do it, I have still requested it. If I am in an appropriate position of authority, I can still hold you responsible for failing to heed my request when you don't, even if I never wanted you to satisfy it in the first place. For some plausible constraints on commanding, requesting, promising, and the like, insincerity is no excuse.

The case of assertion and sincerity is not entirely representative of the class of commitments incurred by a particular illocutionary act. This is because we capture nearly all of the commitment of assertion when we talk in terms of sincerity conditions. Since the purpose of assertion is to convey information about what I believe (about the world) and, typically, to get you to believe it as well, when I am insincere, I typically contravene the usual purpose of my asserting in the first place.²² The close connection between sincerity and commitment also holds for some other illocutionary acts, like that of apologizing. An insincere apology is no apology at all since, presumably, the purpose of an apology is to make manifest my feelings of regret. No feelings of regret, no apology. But it is *not* true for some other common illocutionary acts, like promising, swearing, requesting, demanding and such like. My demand may be insincere, but that doesn't make it fail as a demand—it just makes me a bit of a jerk. And some illocutionary acts do not have sincerity conditions at all—there is no particular mental state required to back up naming one's new cottage Dunroamin.²³

So sincerity conditions are inadequate as a full analysis of the relevant notion of commitment. What we need is a more general picture of commitment on which the commitments of illocutionary acts can spill past what is required for an illocution to be sincere.

3.2 Commitments in General

Sincerity conditions give rise to a particular class of the commitments. Plausibly, they play the role of *constitutive rules* specifying what it is to sincerely ϕ , for ϕ a particular speech act we can engage in (Searle 1969). A sincere promise is one we intend to keep. A sincere assertion is one we believe. A sincere declaration of intention is one where we so-intend. And so on. Not all commitments are connected to sincerity, however. An insincere promise is still a promise, as any 12 year old knows. Likewise an insincere assertion counts as an assertion, as any politician skewered on the Daily Show knows all too well. Sincerity is

²²Note that this is not to say that we can explain what an assertion is merely in terms of this purpose. That is a further step I am not here endorsing. See (MacFarlane 2011) for discussion. See also (Lackey 2007) for useful examples where we can successfully assert while explicitly violating the sincerity condition. This is relevant to §4 below.

²³Or, anyways, no mental state other than the general knowledge of what I am doing. This, of course, is required for nearly all actions, illocutionary or otherwise, for which we can be held responsible.

nice, but it isn't everything.

Why do sincerity conditions give rise to commitments at all? Why sincerely promise? Because there are more than constitutive rules in play in a particular conventional enterprise like the practice of promising. There are also *regulative* rules governing the various practices in play (Searle 1969). Regulative rules, like *drive on the right*, *pay your parking tickets*, or *don't wear white shoes before Easter*, govern existing behavior, (formally) obliging us to act thus and so, at least to those practices we engage in. The reason that sincerity conditions give rise to commitments is that the practices under consideration are governed by the regulative rule: *be sincere!*. Since we have a (formal) obligation to be sincere when we assert, we have an obligation to believe what we assert.

The reason for the parenthetical '(formal)' in the above is that there is often taken to be a distinction between *merely formal* obligations, like the obligation to be polite, and *obligation tout court*, like the obligation to never stab someone just for fun. Merely formal obligations are those which we have to the degree to which we care to engage in a particular conventional activity like that of polite behavior.²⁴ Thankfully, we can bypass this murky dispute in the cases under consideration by noting that the ability to opt out of merely formal obligations is typically muted in cases of illocutionary acts. I can choose not to assert and not to promise, but I cannot simply opt out of the obligations associated with asserting and promising by choosing not to care about them. It's just not that sort of game.

This means that the sanctions which are associated with violating my obligations in the cases of asserting and promising cannot be demurred by saying that I don't care. I'm an involuntary participant in these conventions merely by my participation in the acts of promising and asserting. This is similar to the cases of games in general. I can choose not to play bar pool, but insofar as I play bar pool, I am obliged by the rules tacked on the bar wall. The fact that I don't care about these rules does not excuse me from whatever sanctions are tied to violation of them (typically loss of the game and general disapprobation of my fellow players. And, at least in Boston, getting my name rubbed off the chalkboard.)

3.3 Commitments and Explaining Moore-Paradoxical Utterances

Sincerity conditions are insufficient for explaining Moore paradoxicality since cases like the assertive of swearing to the truth of something, commissives like promising, declaratives like "you're fired" and the like can engender Moore paradoxicality without violating sincerity conditions. The conditions required for the successful carrying out of a swearing, a promising, and a declaration have to

²⁴I am speaking here in the voice of the many; I personally believe that etiquette and like "merely formal obligations" are often given undeserved short shrift. See (Woods forthcoming).

do with the relevant authority to carry out such an action and not with the possession of a particular mental state. As I've said a number of times now, an insincere promise is still a promise and declaring we won't fulfill it is different from declaring that we don't intend to fulfill it. Yet both are Moore paradoxical. Put bluntly, the point is that the sincerity condition of a promise is the possession of an intention, not a performance. Yet we cannot adjoin a declaration of non-performance to a promise without Moore paradoxicality, regardless of what mental state we happen to be in.

If sincerity conditions won't suffice, perhaps the general category of commitments will do. Unfortunately, we here face a problem with EXTENSIONAL ADEQUACY. There are cases in which we explicitly and overtly violate our commitments but where no Moore paradoxicality occurs. Consider, for example, a case in which I promise to give you the original deed to my house. I am absentmindedly lighting a cigarette and do so off of a scrap of paper in my pocket I light off of a campfire. You, hearing my promise, watch as I simultaneously light my cigarette off of the original deed to my house.

Have I committed myself to giving you the original deed to my house? Yes. Have I displayed, overtly and explicitly, my failure to fulfill this commitment? Yes. Has my assertion given rise to a case of Moore's paradox? No. Compare a case far closer to a Moore-paradoxical utterance. We're talking about the deed to my house and it is right there in front of us. I simultaneously set it on fire and promise to give it to you. In such a case, it seems that I have also explicitly failed to fulfill the commitment I'm making, but in such a way as to actually give rise to the Moore-paradoxical feeling. After all, I'm putting right in your face that I am knowingly destroying my ability to fulfill my commitment.

If these behavioral cases don't resonate, we can get a similar effect by means of case of perjury or insincerity. Suppose I—a prominent mafioso—swear to tell the truth, the whole truth, and nothing but the truth. I then swear that I did not murder Tip on the night of July 17th. Immediately released from my oath, I mutter “of course I did, but what do you expect me to say?” Or, alternatively, suppose in the course of a relationship my partner and I are debating about whether we will always care for each other. I, knowing myself, say that I'm almost certain to have lost interest in 5 or 6 years. My partner, desperate for reassurance, begs me to promise him it won't happen. I finally cave and say: “Okay, okay. I promise to love you forever...of course, you and I both know that won't happen.” In such a case, it seems clear to me that I have *still* committed myself to loving him forever; it's just that I will not do so.²⁵

Recapping briefly, we have discussed both a general notion of commitment and the special case of commitments which arise from sincerity conditions. We then saw that the commitments arising from sincerity conditions do not suffice

²⁵The reader is invited to guess at how this example would actually play out in practice.

as an explanation since (a) some cases do not involve sincerity conditions at all, and (b) we can evince our lack of sincerity without giving rise to Moore paradoxicality. Broadening our commitments resolves (a), but it does little to resolve (b) since we can still find cases of Moore paradoxicality even when we have explicitly violated our commitments. Something else is needed for commitments to explain our target phenomenon.

4 Commitment-Theory Done Right

If commitment and explicit demurral thereof is insufficient to explain Moore's paradox, then what needs to be added? We find the answer to this in a common thought about illocutionary actions, articulated nicely in principle form by Bach and Harnish:

COMMUNICATIVE PRESUMPTION The mutual belief in the linguistic community C_L to the effect that whenever a member S says something to another member H , S is doing so with some recognizable illocutionary intent. (Bach and Harnish, 1979: 60-61)

Bach and Harnish take this presumption, held at least implicitly by all competent speakers and interpreters, to govern successful interpretation of the speech acts of others. Their focus is on communicative acts and, in particular, on the expression of sincerity conditions in carrying out these communicative acts. They take expression of a sincerity condition to be governed by the scheme:

For S to *express* an attitude is for S to R -intend the hearer to take S 's utterance as reason to think S has that attitude. (Bach and Harnish, 1979: 15)

and to R -intend (reflexively intend) something is to intend that the intention to do it be recognized (typically by the interpreter.) Putting these together, when someone says something with recognizable illocutionary—here expressive—intent, we take them to intend that our recognition of their intention gives us reasons to believe that they believe what they say; we are thereby generally entitled to attribute to them belief in what they have asserted.

This picture is geared towards a communicative picture of illocutionary acts. Given our focus on cases where the illocutionary act, such as a promise, gives rise to a commitment to *act*, not merely to *intend to act*, we need to look elsewhere than expression to explain how to get Moore's paradox out of commitments. (See also §3.1 above.) Luckily, we don't have to look far. Expression is one thing that we can do with illocutionary intent. Another is to take on commitments incurred by participating in an illocutionary practice. We can lay this down by means of a schema, as above:

For S to *commit themselves* by means of an utterance is for S to R -intend the hearer to take S 's utterance as reason to think S is

voluntarily engaging in the practice and thereby taking on the commitments engendered thereby.

Putting this together with COMMUNICATIVE PRESUPPOSITION, we obtain that when someone says something with recognizable illocutionary intent, we take them to intend that our recognition of their intention gives us reasons to think *S* is voluntarily engaging in that practice; we are thereby generally entitled to treat them as recognizing that they are thus and so committed and acting accordingly.

If we tie this into the above discussion of commitments, we can see that treating an utterer as recognizing that they are thus and so committed entitles us to treat them as being *bound* and *acting in accord*s with their commitments. At least those that are immediately recognizable. We take those who recognize their commitments to also intend to satisfy those commitments any linguistically competent ϕ -er, where ϕ is a speech act like asserting, will immediately recognize.²⁶ When they go on to explicitly deny that they will meet these commitments, by denying they'll satisfy their commitments, we obtain the sort of self-undermining that characterizes Moore paradoxicality. If they recognize that their voluntary engagement with this practice commits them and requires taking these commitments as guiding their action, why would they deny that they'll satisfy these commitments?

Let's work through a particular case. Suppose you promise to give me 5 Turkish lira, but then immediately deny that you will do so. By COMMUNICATIVE PRESUPPOSITION, I take your utterance to be knowably intended to be a promise. I also take promising to be a way of committing yourself to keeping the content of the promise. Since you are committing yourself, I take you to intend me to recognize your intention to subject yourself to the commitment to keep your promise. But, given that I recognize this and recognize that you want me to recognize it, I am completely puzzled as to why you would immediately go on to disabuse me of my recognition of your intention to submit yourself to the commitment to keep your promise. Crash.

Of course, this does not mean that someone cannot insincerely engage in a practice or, after so engaging, lose their interest or ability to satisfy their intention to keep their commitments. Of course they can. This explains why the third-personal, past-tense, and otherwise non-first-personal cases of Moore's paradox aren't paradoxical. There are lots of reasons for taking someone to no longer intend to satisfy their commitments. What's important here is that the first-personal acts involving committing oneself, like the first-personal acts involving expressing, gives rise to a presumption that we *then* intend to keep our commitments. Frustration of this presumption causes the paradoxical feeling.

²⁶This hedge is necessary to explain why far-out violations of commitments can be coherently tacked onto something like a promise. I might just not recognize that I cannot both have lunch with you and pick up Bob at Atatürk airport within 5 hours. Further discussion of this obviously and easily solvable problem would distract, not illuminate.

When this presumption is not in force, such as when there is a contextual reason to interpret our illocutionary action non-standardly, as in the cases described in the last section, then we do not have any paradoxical feeling. This is exactly the right result.²⁷

Essentially, this theory ties together a commitment-theoretic picture of (some) illocutionary acts²⁸ with a Bach/Harnishian account of how to interpret speech-acts. This allows us to retain the idea that many illocutionary acts commit us to things other than the possession of mental states, but to also explain why, in typical occurrences of illocutionary acts, we are permitted to attribute to the utterer a particular sort of intention: an intention for us to recognize their intention to subjugate themselves to the norms of the (conventions governing) particular illocutionary acts they have engaged in. This view has the advantages of both the commitment view—it is sufficiently divorced from the theory of verbal communication so as to allow the cases of behavioral Moore’s paradox like the one above involving flagrantly lighting the deed to house on fire—while retaining the advantageous non-mysteriousness of the Bach/Harnish communicative presupposition account. It allows us to satisfy UNIFORMITY as the picture captures all cases of Moore paradoxicality, EGALITARIANISM since the main assumption is that competent speakers and interpreters presume illocutionary intent, and has at least the potential to satisfy EXTENSIONAL ADEQUACY as we can explain why not all commitments incurred by speech acts give rise to Moore’s paradox.

4.1 External Conventions, Attributed Intentions

One worry with the above account is that it presumes that speakers commit themselves by actually intending to carry out their illocutionary act. This may strike some—it strikes me as possibly—psychologically unrealistic. It also seems to neglect the fact that we engage in speech acts without thinking through what we are accomplishing by means of them. We all know *on reflection* that by carrying out various illocutionary acts we commit ourselves thus and so. However, it is nowise clear that this is clear to us, explicitly, at the time that we are actually carrying them out.

If this worries you, I will note that I am torn between the above account of commitment and more externalist alternatives on which:

For *S* to *commit themselves* by means of an utterance is for *S* to be fairly attributed the *R*-intention for the hearer to take *S*’s utterance

²⁷We can resolve other problems for utterances with Moore-paradoxical form, like those in (Atlas 2007), by noting that context can suggest that we interpret reflexive pronouns, like ‘I’, in non-standard ways in special contexts like those of an amnesiac going over “his” past with a therapist.

²⁸Compare with the commitment-theoretic picture of assertions discussed in (MacFarlane 2011).

as reason to think S is voluntarily engaging in the practice and thereby taking on the commitments engendered thereby.

or the combination of the even more externalist schema:

For S to *commit themselves* by means of an utterance is for S 's utterance to subsume them under a norm governing that type of utterance.

and the principle:

Whenever a member S says U to another member H , absent special conditions, S should be interpreted as R -intentionally committing themselves by means of U .

Any of these views would give a recipe for satisfying UNIFORMITY, EGALITARIANISM, and potentially EXTENSIONAL ADEQUACY. Which is actually correct is a matter for another time and another paper. Note also that I have spent fairly little time talking about what norms actually govern various illocutionary acts. In the case of assertion, this has been extensively discussed (see (Weiner 2007) for a useful overview) and in the general case of speech acts, a reader can find useful accounts in both (Searle and Vanderveken 1985) and (Bach and Harnish 1979). Again, little turns on this for my point, so long as my above assumptions about particular speech acts and their commitments are close enough. All I need is that some speech acts commit us to things other than possessing various mental states in order to motivate my general picture; which things we are actually committed to is mostly besides the point for this.

5 Conclusion

I have argued that in order to accommodate the wide variety of Moore paradoxical phenomena, we need to abandon the idea that we can explain Moore paradoxicality in speech by means of Moore paradoxicality in belief. I went on to argue that we could not use irrationality in speech or insincerity to run this sort of explanation; rather, we need to move to a commitment-theoretic account of Moore paradoxicality. I then sketched the details of this view, arguing that we needed a presumption about the intentions of speakers (adopted, in modified form, from Bach and Harnish) to separate the commitments incurred by speech acts which gave rise to Moore paradoxical reactions and those which did not. The resulting picture vindicates UNIFORMITY, EGALITARIANISM, and potentially vindicates EXTENSIONAL ADEQUACY. (subject to a suitable filling in of the gaps.)

I will close on a final advantage of this kind of view. I suggested above (and in my (2014)) that the known irrational can still make promises, declarations, and assertions in ways that give rise to Moore paradoxicality. Suppose we make the presumption, suggested above, that when someone typically says something

to someone else, they are doing so with illocutionary intent. If we presume this for ordinary folk, presumably we should also presume it of the irrational as well *unless* we know that their irrationality is manifesting in lack of ability to successfully communicate. After all, irrationality is not an all or nothing affair and even the raving mad can sometimes successfully request a glass of water. If we presume this sort of interpretive charity in assessing others, then it seems that we should sometimes be startled when even someone known to be irrational deliberately undercuts themselves in requesting, then denying that they want something. We can back-solve to see that they weren't really voluntarily engaging in an illocutionary act (or, alternatively, deliberately flouting their commitments), of course. But no one, to my knowledge, has ever successfully argued that Moore-paradoxical utterances can't be later explained. It is their current bizarreness that is so remarkable.

References

- Alston, W. 2000. *Illocutionary Acts and Sentence Meaning*. Ithaca and London: Cornell University Press.
- Atlas, J. D. 2007. What Reflexive Pronouns Tell Us About Belief : A New Moore's Paradox de Se, Rationality, and Privileged Access. In *Moore's Paradox: New Essays on Belief, Rationality, and the First Person*, ed. M. S. G. a. J. N. Williams, 117–145. New York: Oxford University Press.
- Bach, K. and R. Harnish. 1979. *Linguistic Communication and Speech Acts*. Cambridge, Mass. and London: MIT Press.
- Baker, D. 2014. The Abductive Case for Humeanism Over Quasi-Perceptual Theories of Desire. *Journal of Ethics and Social Philosophy* 8(2): 1–29.
- Baker, D. 2015. Akrasia and the Problem of the Unity of Reason. *Ratio* 28(1): 65–80.
- Baker, D. and J. Woods. 2015. How Expressivists Can and Should Explain Inconsistency. *Ethics* 125(2): 391–424.
- Baldwin, T. 2007. The Normative Character of Belief. In *Moore's Paradox: New Essays on Belief, Rationality, and the First Person*, ed. M. S. G. a. J. N. Williams, 76–89. New York: Oxford University Press.
- Brandom, R. 1983. Asserting. *Noûs* 17(4): 637–650.
- Brandom, R. 1994. *Making It Explicit: Reasoning, Representing, and Discursive Commitment*. Cambridge, Ma: Harvard University Press.
- Harnish, R. M. 2005. Commitments and Speech Acts. *Philosophica* 75.
- Heal, J. 1976. Insincerity and Commands. *Proceedings of the Aristotelian Society* 77: 183–201.
- Lackey, J. 2007. Norms of Assertion. *Noûs* 41(4): 594–626.
- MacFarlane, J. 2005. Making Sense of Relative Truth. *Proceedings of the Aristotelian Society* 105(3): 321–339.

- MacFarlane, J. 2011. What Is Assertion? In *Assertion*, eds. J. Brown and H. Cappelen. OUP Oxford.
- Michaelson, E. 2016. The Lying Test. *Mind and Language* 31(4): 470–499.
- Moore, G. 2013. *G.E. Moore: Selected Writings*. London: Routledge.
- Moore, G. E. 1942. A Reply to My Critics. In *The Philosophy of G. E. Moore*, ed. P. A. Schilpp. La Salle, IL: Open Court.
- Moran, R. A. 2001. *Authority and Estrangement: An Essay on Self-Knowledge*. Princeton: Princeton University Press.
- Raskoff, S. manuscript. Getting Expressivism out of the Woods.
- Searle, J. 1969. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.
- Searle, J. and D. Vanderveken. 1985. *Foundations of Illocutionary Logic*. Cambridge: Cambridge University Press.
- Shoemaker, S. 1995. Moore’s Paradox and Self-Knowledge. *Philosophical Studies* 77(2): 211–228.
- Sorensen, R. A. 1988. *Blindspots*. Oxford: Clarendon Press.
- Stainton, R. J. 2016. Full-On Stating. *Mind and Language* 31(4): 395–413.
- Toppinen, T. 2014. Expressivism and Moore’s Paradox: A Reply to Woods. *Journal of Ethics and Social Philosophy*: 1–5.
- Toppinen, T. 2015. Relational Expressivism and Moore’s Paradox. *Journal of Ethics and Social Philosophy*: 1–7.
- Weiner, M. 2007. Norms of Assertion. *Philosophy Compass* 2(2): 187–195.
- Williams, J. N. 2012. Moore-Paradoxical Assertion, Fully Conscious Belief and the Transparency of Belief. *Acta Analytica* 27(1): 9–12.
- Williams, J. N. 2013. Moore’s Paradox and the Priority of Belief Thesis. *Philosophical Studies* 165(3): 1117–1138.
- Williams, J. N. 2015. Moore’s Paradox in Speech: A Critical Survey. *Philosophy Compass* 10(1): 10–23.
- Woods, J. 2014. Expressivism and Moore’s Paradox. *Philosophers’ Imprint* 14(5): 1–12.
- Woods, J. 2015. Expressivism Worth the Name: A Reply to Teemu Toppinen. *Journal of Ethics and Social Philosophy*: 1–6.
- Woods, J. forthcoming. The Normative Force of Promising. *Oxford Studies in Normative Ethics* Vol. 6.
- Woods, J. manuscript committed to a volume. The Authority of Formality. *Oxford Studies in Metaethics*.
- Wright, C. 1992. *Truth and Objectivity*. Cambridge, Ma: Harvard University Press.