Philosophia (2019) 47:313–322
https://doi.org/10.1007/s11406-018-0004-8

Suspicious Minds: Coliva on Moore's Paradox and Commitment

Aidan McGlynn[1]

Published online: 26 July 2018

1 Introduction: Varieties of Self-Knowledge and Varieties of Belief

In *The Varieties of Self-Knowledge*, Annalisa Coliva defends a pluralist approach, according to which we know about our different kinds of mental states in different ways. The case for pluralism in some area of philosophy is often rather defeatist in spirit. We try out a number of different explanations for some phenomena, and realise that each falls short; at best each explanation looks good for some initial, restricted diet of examples, but once we consider the full range of the phenomena to be explained, things go awry (e.g. Wright 2013: 123). This kind of reasoning to a pluralist conclusion might plausibly apply to the debate on self-knowledge, where there's been a tendency to construct accounts that perhaps work well enough for either beliefs, or phenomenal states such as pains, but struggle badly when applied more broadly. Coliva's book isn't free of this kind of reasoning (and there's no reason it should be!), but there's also a more positive, bottom-up case for pluralism: an argument that takes seriously the variety between the states supposedly self-known, and posits an appropriate variety of ways that states can be self-known.

Interestingly, Coliva's positive case for pluralism can be made without considering any other mental states beyond belief, since she argues we need a distinction between *beliefs as dispositions* and *beliefs as commitments*, and this already forces any comprehensive account of self-knowledge to be heterogeneous. Beliefs as dispositions, as characterised by Coliva, are 'not the result of conscious deliberation' or the assessment of evidence, not under one's direct control, and not states that one can be held 'rationally responsible' for (2016: 28), while beliefs as commitments 'depend on a judgement based on the *assessment* of the evidence at subjects' disposal and that, for this reason, are within their control and for which they are held rationally responsible' (2016: 32). While beliefs as commitments can be self-known in a distinctively first-personal way, we can at best have 'third-person' self-knowledge of our beliefs as dispositions; we can know about them though self-directed mindreading, involving

✉ Aidan McGlynn
amcglynn@exseed.ed.ac.uk

[1] School of Philosophy, Psychology, and Language Sciences, The University of Edinburgh, Edinburgh, UK

interpretation of our own behaviour (verbal and non-verbal) plus 'other mental states one has knowledge of in a first-personal way, such as one's sensations, immediate feelings and propositional attitudes as commitments' (2016: 70).

Why think we have two species of *belief* here (rather than, for example, belief proper and something like Tamar Gendler's notion of an *alief*: see Gendler 2008)? To the extent that we use the same word for both, why don't the differences gestured at in the previous paragraph suggest the situation with 'belief' is much like the situation with 'jade'? Coliva responses that there is 'considerable overlap…to justify the idea that they are more like two species of the same genus, and therefore it is no mere accident that the term "belief" is, in fact, applied to both' (2016: 37). Coliva doesn't say much about what this overlap consists in, but she does suggest that when one has a belief as commitment that P, one has 'at least some of the same behavioural dispositions' one would have if one had a belief as disposition with the same content'.

2 Moore's Paradox

There's a lot more that could be said about this distinction within the class of beliefs, and the role that it plays in Coliva's pluralist theory of self-knowledge, but my attention here will be on its role in her discussion of Moore's paradox in the appendix to the book. As it is usually thought of, Moore's paradox is that assertions of the form 'P, but I don't believe that P' sound contradictory even though there's no contradiction between the two conjuncts of what I assert. In contrast to any explicit contradiction, the past tense ('P, but I didn't believe that P') and third person analogues of Moorean conjunctions ('P, but she didn't believe that P') seem totally unremarkable; likewise, there's no problem with *supposing* rather than asserting the original conjunction.

The boundaries of the paradox are disputed, but it's relatively uncontroversial that it extends along a number of dimensions. First, as Coliva notes, John Williams (1979) has pointed out that Moore offers examples showing that claims of the form 'I believe that P, but not-P' seem equally bad; Williams calls the original the *omissive* version, and this variant the *commissive* version. Second, Moore also suggested that assertions of the form 'P, but I don't know that P' are also paradoxical; elsewhere (McGlynn 2013), I've called these *epistemic* Moorean assertions, to distinguish them from the original *doxastic* examples.[1] Third, at least since Roy Sorensen's *Blindspots* (1988), it's been widely recognised that *beliefs* of the form 'P, but I don't believe P and 'I believe that P, but not-P' seem just as paradoxical as the corresponding assertions; despite the contents of such beliefs being consistent, believing a doxastic Moorean conjunction seems to be deeply irrational or absurd.

There are a number of different accounts of Moore's paradox, offering competing explanations of how believing or asserting a perfectly consistent content can give rise to contradictoriness, absurdity, or irrationality. Moreover, as Coliva notes (Coliva 2016: 243–4), many of these accounts encourage us to draw significant morals from the paradox. These include morals concerning self-knowledge and de se thought and talk, the norms constitutively governing assertion and belief, and so on. In Coliva's appendix, she offers a diagnosis of what's going on in the paradox, focusing on (in the

[1] There's no commissive version of the epistemic paradox, given the factivity of knowledge.

terminology I've introduced above) doxastic Moorean beliefs, in both their omissive and commissive versions, though Coliva also touches on issues around Moorean assertions. Moreover, she too wants to draw a significant moral, namely that 'any satisfactory account of Moore's paradox will have to unravel the complexity of our concept of belief' (2016: 255). Accounts that don't appeal centrally to the distinction between beliefs as commitments and beliefs as dispositions will 'lose' the paradox (in a sense I'll say more about shortly); indeed, we not only need the notion of belief as commitment to save the paradox, but 'we need to stick to a rather resolute notion of commitment' (2016: 244). Let's unpack what Coliva might mean by this, and how she argues for this moral.

Coliva focuses on the following two forms of Moorean conjunction:

(1) I believe that P, but it is not the case that P
(2) I do not believe that P, but it is the case that P

These are the commissive and omissive versions of the paradox respectively, with the order of the conjuncts swapped from my original presentation above. Since I don't think the order of the conjuncts matters, let's work with these forms instead of the usual ones.[2] Now, before evaluating explanations of the absurdity of assertions of things like (1) and (2), and the apparent absurdity or self-defeat involved in believing either conjunction, Coliva asks us to consider an example that motivates some constraints on what a response to the paradox should look like, as well as the 'diagnostic suggestion' that points the way to Coliva's preferred response. Here's the example:

Jane's Odd Case

Jane is married to Jim. They have been married for several years and have a daughter. Jane is often at home, on her own, attending to domestic chores. From time to time, she feels lonely and wishes that she had pursued her own career. More often than not, however, she feels much rewarded by the fact that her family is so serene. Indeed, when she meets up with her friends, who sometimes complain about their husbands, she cannot help remarking that her life makes her happy and that her husband is adorable and complete trustworthy.

Still, it often happens that, while preparing for the laundry, Jane carefully searches Jim's pockets. While tidying up his studio, she opens and examines all the drawers. While dusting the furniture, she lingers on the screen of his laptop, left open on the incoming messages.

One day, Ann, a psychoanalyst friend of Jane's, approaches her and tells her about Freud's theories concerning the unconscious. Little by little, the deep significance of a whole series of previously meaningless actions is disclosed to Jane. Ashamedly, she realised that all that attention spent over the content of her husband's pockets was a sign of her being insecure about him. All that dusting the screen of his laptop, a symptom of her thinking that he might have some intimate

[2] Not everyone agrees that the order is irrelevant: see e.g. Gilles 2001: 247–8.

> correspondence with another woman. Still, Jane knew all too well that Jim had always been the most truthful of men. The thought popped into her head: "I do believe that Jim is unfaithful to me, but he is not". (2016: 253)

Coliva takes the example to show that judgments of the form 'I believe P, but not-P' can be 'perfectly legitimate', and this raises the threat that 'Moore's paradox is *dissolved*' (2016: 254, italics in original). Why might the paradox be dissolved? Coliva isn't explicit, but the natural interpretation is that the paradox (at the level of thought) is standardly presented, as it was above, as that associated with the absurdity or self-defeat involved in believing things of the form of (1) and (2), despite the fact that such conjunctions will often be manifestly consistent. If one can believe something that has the form of (1) but without absurdity or self-defeat, precisely the possibility that Coliva takes Jane to illustrate, then that seems to undermine the central claim underlying the paradox. Attempts to explain why doxastic Moorean beliefs are invariably problematic will simply be misguided.

However, Coliva doesn't think that the paradox is so easily vanished (nor that it would be a good thing if it were, given its philosophical significance). Instead, she articulates some conditions which spell out when beliefs in things of the form of (1) and (2) will generate the paradox. In doing so, Coliva also offers a constraint on possible treatments of the paradox, namely that they shouldn't overgeneralise to cases (including Jane's odd case) in which her conditions aren't met. Coliva also thinks that the example motivates the 'diagnostic suggestion' that 'any satisfactory account of Moore's paradox will have to unravel the complexity of our concept of belief' (2016: 255). As noted in the introduction, this doesn't just mean that she thinks that such we will need to appeal to the distinction between beliefs as dispositions and beliefs as commitments, but that 'we need to stick to a rather resolute notion of commitment' (2016: 244).

3 Coliva's Account

To begin, we need to table Coliva's conditions for when a doxastic Moorean belief—a belief in something of the form of (1) or (2)—will be genuinely Moore paradoxical. First, one's first-order belief must be an item of self-knowledge: something that one is conscious of (2016: 255). In Jane's odd case, this constraint is met after, but not before, she becomes aware that she believes that her husband is being unfaithful to her. As this brings out, for a mental state to be self-known or conscious to one in this sense doesn't require that one acquire that knowledge or consciousness in a distinctively first-personal way; above we saw that Coliva recognises a category of 'third-personal self-knowledge', and as I understand her, this is all that her first condition requires. Additionally, the belief being attributed to one in (1) must be a belief as commitment rather than a mere disposition. It's this condition that fails to be met in Jane's odd case even after she recognises her belief that her husband is unfaithful; when she comes to believe something of the form of (1), she is merely attributing to herself a belief as disposition, with the relevant disposition being revealed in her jealous behaviour. That's why her belief, despite being Moorean in form, is not Moore paradoxical. Coliva's proposal is that these conditions serve to pick out a subset

of beliefs in Moorean conjunctions that are genuinely paradoxical, while leaving room for cases such as Jane. The paradox may shrink a little, but it doesn't disappear.

In 'saving' the paradox in this way, we've already had to invoke the distinction between beliefs as dispositions and beliefs as commitments. We haven't as yet seen any need for the 'rather resolute' notion of commitment that Coliva takes to be necessary to do justice to the paradox, but we are in a position to ask the question that will bring it into view: why are doxastic Moorean beliefs that meet Coliva's two conditions inherently irrational or self-defeating?

Coliva offers two responses to this question, with the second improving on the first, by Coliva's lights, precisely because it invokes a resolute notion of commitment. Since Coliva's second response builds on her first, let's take them in turn. Despite there being different accounts of what having a belief as commitment requires, Coliva takes these accounts to all agree that 'beliefs as commitments are attitudes of acceptance of a given propositional content that are intrinsically normative' (2016: 257). What does 'intrinsically normative' mean here? There are weaker and stronger answers to this question invoked in Coliva's discussion. It's the weaker answer that's relevant to Coliva's first treatment of Moore's paradox, as far as I can tell; we'll come back to the stronger answer when we consider her 'second pass' at the paradox below.

Here's Coliva's 'first pass'. According to the weaker answer, beliefs as commitments are 'intrinsically' normative in the sense that having such a belief that P involves seeing that one ought to use P as a premise in one's (practical and theoretical) reasoning. Now, consider (1) again:

(1) I believe that P, but it is not the case that P

The idea is that believing this would involve a certain kind of irrationality. Suppose, in line with Coliva's conditions on when a belief of this form is genuinely Moore-paradoxical, that the first conjunct picks out a belief as commitment. Then I'm self-ascribing a belief that requires me to recognise that I ought to use P as a premise in my reasoning. But in asserting that it's not the case that P, I thereby commit to using the negation of P as a premise in my reasoning, and I'd thereby be guilty of irrationality:

> Thus [one] would commit [oneself] (knowingly and willingly) to reasoning from contradictory premises. And *this* would be irrational. (2016: 260, italics in original)

Coliva's first treatment of (2) is similar; I won't go through the details for considerations of space.

However, Coliva isn't satisfied with this explanation of what's wrong with doxastic Moorean beliefs (in cases where her conditions are met). Here's her worry. Suppose that one can 'knowingly and willingly' reason from contradictory premises, despite the irrationality of doing so. Then in believing (or asserting) something with the form of (1) in such a case, one would just be correctly representing on one's own irrational state of mind: 'if it were possible for a subject to endorse incompatible commitments, what would then prevent one from judging and even from asserting that one did?' (2016: 261). So again, the paradox seems to go missing; instead of explaining why believing (1)

would itself be irrational, this first account leaves open the possibility that believing (1) might be a perfectly correct and reasonable representation of an irrational state of mind.

Doing better involves the advertised 'resolute' notion of commitment. Coliva's 'first pass' ran aground on the possibility that one might knowingly and willingly reason from contradictory premises, and so the idea is to close off that possibility: we need to understand the notion of commitment in 'a more demanding way and, in particular, as entailing a negative answer to the question whether it is possible knowingly and willingly to hold incompatible commitments' (2016: 261–2). Here Coliva appeals to the stronger interpretation of the claim that beliefs as commitments are 'intrinsically normative', mentioned above. Coliva quotes the following passage from Bilgrami to elaborate:

> In this latter normative usage, to desire something, to believe something, is to think that one ought to do or think various things, those things that are entailed by those desires and beliefs by the light of certain normative principles of inference (those codifying deductive rationality, decision-theoretic rationality, perhaps inductive rationality and also perhaps some broader forms of material inference having to do with the meanings of words as well). (2006: 213)[3]

On Bilgrami's view, one doesn't need to in fact do the things entailed by one's attitudes, given these (often implicitly grasped) normative principles, but one does need to think one ought to do them. Coliva is more neutral on whether the latter by itself suffices for one to count has having the relevant commitments (2016: 259). What's important for her purposes is instead what she takes to be an implication of this way of developing the idea that beliefs as commitments are 'intrinsically normative', namely that having a belief as commitment involves 'seeing oneself as having to implement a certain sort of behaviour', and that 'there is an internal link between the content of one's belief as commitment and the kinds of actions that one ought to perform given that belief' (2016: 258–9). I find this a bit puzzling. We can readily come up with plausible normative consequences of someone's having a belief as commitment that P which are intimately tied to the content of that belief; for example, perhaps to have this belief as commitment involves recognition that one ought to use P as a premise in one's theoretical and practical reasoning, or recognition that if someone asks one whether one believes that P one ought to, other things being equal, answer in the affirmative, and so on. These are recognisably versions of the weaker interpretation of the claim that beliefs as commitments are intrinsically normative distinguished above, and Coliva sometimes writes as if this is all she has in mind (e.g. 2016: 259–60).[4] But the passage we've just quoted, we get the idea that there are *specific* 'courses of action' that one must see oneself as committed to undertaking, and which can potentially be specified in independent terms, but which are still internally linked to the content P. Here's Coliva's own example:

> For example, if I believe that it is raining and I do not want to get wet, I ought, *ceteris paribus*, to go out with an umbrella, whereas if I merely so wished, I would not. (2016: 257)

[3] Coliva mistakenly references this passage to Bilgrami 2012.

[4] Thanks to discussion with Daniel Stoljar here.

This helps to illustrate what I said in the previous paragraph. Going out with an umbrella is a specific course of action, and it's one specified in terms distinct to those used to specify the content of the belief that it is raining. Of course, there is a relationship between the relevant terms here (given that an umbrella just is a particular kind of device with the function of protecting its user against rain), and recall from the passage from Bilgrami quoted above that the commitments one needs to recognise might include those entailed by one's attitudes in light of 'some broader forms of material inference having to do with the meanings of words as well'. But this semantic link acknowledged, I'm not sure in what sense there's an 'internal link' between the content of the belief that it is raining and the course of action of taking an umbrella. Notice that in Coliva's example, it's specified that one has a particular desire, not getting rained on, and there's a ceteris paribus clause in the mix too; this clause will presumably require that it's the case that there's no other equally good or better way to satisfy my desire to stay dry, given my belief that it's raining, and it may gesture at a number of other conditions too. Given all this, it's not clear to me in what sense there's an 'internal link' here. Taking P as a premise in one's practical reasoning is plausibly a normative commitment that has an internal link to the content of one's belief that P; I'm less clear how there could be an internal link between P and particular courses of action that would be the conclusion of such practical reasoning, were one's desires a certain way and all other things equal.

Having registered this point of puzzlement, let me continue to lay out Coliva's second and final treatment of Moore's paradox. The idea, as I understand it, is to appeal to the internal link between the contents of beliefs as commitments and the associated mandated courses of action in order to argue that knowingly and willingly holding inconsistent beliefs as commitments is impossible. Why would this be impossible? Because 'the courses of action mandated by inconsistent commitments are mutually exclusive', such that to commit oneself to the course of action mandated by the negation of P is to thereby relinquish any commitment one had to the course of action mandated by P, and vice versa; one *cannot* be committed to both at the same time (2016: 262). Coliva again appeals to her example to illustrate this. Suppose Jane wants to remain dry, and the only way she can do so is to open her umbrella.

> Now, if she has the belief as a commitment that it is raining, she ought to see herself as bound to opening the umbrella. If she has the belief as a commitment that it is not raining, she ought to see herself as bound to not opening the umbrella. Finally, if she is open minded with respect to whether it is raining, she ought to see herself as bound neither to opening the umbrella nor to keeping it close[d]. The three kinds of action, which are respectively internally linked to each kind of belief as a commitment a subject might have, however, are mutually exclusive.' (2016: 262)

Since having a belief as commitment involves recognising the course of action it 'internally mandates', and since the courses of action mandated by inconsistent premises are mutually incompatible, it is *impossible* to bind oneself to the commitments of inconsistent premises. But that's precisely what would be involved in believing something of the form of (1), in circumstances where Coliva's conditions for genuine instances of Moore's paradox to arise are met, and so no such belief is possible (and

there's a similar story to be told about belief in conjuncts of the form of (2)). Moreover, asserting things like (1) or (2) 'would be self-defeating in the sense of expressing an impossible cognitive situation' (2016: 254).

I find this treatment of the paradox rather complicated, and as I've already mentioned, I'm puzzled by the idea that certain beliefs (as commitments) 'internally mandate' certain courses of action in virtue of their contents. Practical reasoning to conclusions about what course of action one ought to taken, given that one has a particular belief, seems like a messy business—one which depends on what else one believes, one's other relevant desires and intentions, and whatever other factors need to be covered by the ceteris paribus clause that Coliva herself makes use of. Given this, I'm not sure in what sense a particular course of action might be 'internally mandated' by a particular belief as commitment. Further, Coliva's conclusion, that beliefs in things of the form of (1) and (2) which meet her other two conditions are *impossible*, strikes me as rather too strong; such beliefs would be deeply irrational and self-defeating, and this needs explanation, but I find the claim that they are impossible surprising. As a corollary of this, I have a worry with Coliva's proposed explanation of the absurdity/contradictory-feel of doxastic Moorean assertions. As we saw a moment ago, her idea is that such assertions 'would be self-defeating in the sense of expressing an impossible cognitive situation'. That might explain why we ought not assert things like (1) and (2). But why do we *hear* such assertions as problematic? The answer 'because they express an impossible cognitive situation' only seems to hold water to the extent that this is something hearers are, implicitly at least, aware of. To the extent that my own reaction—that it would be surprising were doxastic Moorean beliefs (meeting Coliva's conditions) to prove impossible—is not idiosyncratic, this explanation of why we hear the assertions as absurd seems dubious, even if Coliva is correct that such beliefs would be impossible.

4 Less is Moore

In other work (McGlynn 2013), I've favoured a comparatively minimalistic account of doxastic versions of Moore's paradox, defended in various places by John Williams (e.g. Williams 1996, 1998; Green and Williams 2007: 10–11). The key claims of the account are that belief distributes over conjunction, and that we're sensitive to this fact. Suppose that I form a belief of the form (2): 'I do not believe that P, but P'. Since belief distributes over the conjunction, it follows that I believe that P, and this is a consequence I recognise. But clearly my belief can only be true if I don't believe that P. So the content of my Moorean belief is false, if I believe it; and this is something that only a little reflection will let me appreciate. In this sense, any belief of this form would be inherently and obviously self-defeating. This account can be extended to the commissive version of the paradox, (1), though with some differences. Suppose I believe something of the form 'I believe that P, but it is not the case that P'. Since belief distributes, it follows that I believe P. The truth of the content of my belief, however, requires that I believe the negation of P. So if I have this belief, and it's true, it follows that I believe P and I believe its negation; this belief would require me to be patently inconsistent. And once again, this line of

reasoning is transparent to me, meaning that such a belief is also rationally self-defeating in a way.[5] We can then say a number of things about doxastic Moorean assertions. Perhaps the most natural is a straightforward analogue of Coliva's own account of such assertions; such assertions are attempts to express inherently and obviously self-defeating states of mind.

This style of explanation of doxastic Moorean beliefs (and assertions) is pleasingly simple and economical. It's not without assumptions and consequences that can be challenged, of course. One might dispute the principle of doxastic logic it relies on, or the further claim that this is a principle that figures in our thinking about ourselves and others. One might be concerned about the asymmetric treatments of the omissive and commissive variants of the paradox. One might worry that the explanation cannot be readily extended to beliefs of the form *P, but I don't know that P* (though personally I think that this is an advantage: see McGlynn 2013). But if one accepts this kind of approach to Moore's paradox, it's likely one will be wary of claims that reflection on the paradox has significant pay-offs when it comes to understanding agency, self-knowledge, the nature and norms of belief, and so on.

The question now is whether a style of explanation this thin might be adequate. It does seem to do better than some of the other rival explanations criticised by Coliva in her appendix. Like her own account, it takes the paradox at the level of thought to be primary, rather than treating the paradox as an issue at the level of pragmatics and speech act theory, and it doesn't have the problematic implication that doxastic Moorean conjunctions are disguised contradictions. What about Jane's odd case, though?

Recall that Jane self-attributes a belief that she'd express with 'I do believe that Jim is unfaithful to me, but he is not', based on inference from both her jealous behavior and her knowledge of Jim's fidelity. A straightforward application of the account just offered of (1) seems to yield the result that this belief is inherently and obviously self-defeating, and this conflicts with Coliva's verdict that the belief is 'perfectly legitimate'. However, I'm not totally convinced that Coliva's verdict here is right. No doubt the long backstory reduces the oddness or absurdity of Jane's belief and assertion somewhat. However, it doesn't remove it entirely, as I think we can draw out by comparing her belief and assertion to a variant in the past tense:

'I *do* believe that Jim is unfaithful to me, but he is not'

'I *did* believe that Jim is unfaithful to me, but he is not'

Even with Coliva's detailed backstory for Jane's predicament in mind, an assertion of the second of these conjunctions seems devoid of the characteristic absurdity of Moorean assertions in a way that, to my mind, contrasts with Jane's actual assertion

[5] Objecting to a different kind of approach to the paradox, Coliva notes that a conjunction of the form 'I believe that P and I believe that not-P' is 'not a contradiction or—as such—an absurd judgement' (2016: 251). But on the account I'm recommending, it's not this conjunction that's absurd, but rather knowingly forming a belief that can only be true if one has such patently contradictory beliefs. Moreover, the principle at work in the account in the text is a relatively uncontroversial basic tenant of doxastic logic—belief distributes over conjunction—while the account criticised by Coliva invokes Gareth Evans's elusive and controversial transparency thesis.

of the first. Likewise, a belief in the claim expressed by the latter seems much less absurd or irrational than Jane's belief in the case as described by Coliva. The backstory lessens the perceived absurdity or irrationality of Jane's belief and assertion (as detailed backstories are quite generally apt to do) but the paradox doesn't 'disappear'.[6]

In this final section, I've briefly tabled an account of the absurdity of doxastic Moorean beliefs (and the corresponding assertions) which doesn't require Coliva's 'resolute' notion of belief as commitment—indeed, it's an account that, as I've developed it, is rather insensitive to the distinction between beliefs as dispositions and beliefs as commitments (though I don't mean to deny that such a distinction can or should be drawn). If this account of the paradox is viable, it stands as a challenge to Coliva's 'diagnostic suggestion' that 'any satisfactory account of Moore's paradox will have to unravel the complexity of our concept of belief'. I predict that Coliva will not be persuaded that it takes Moore's paradox seriously enough; for my own part, I continue to suspect that philosophers are inclined to take the paradox a bit too seriously.[7]

References

Bilgrami, A. (2006). *Self-knowledge and resentment*. Cambridge: Harvard University Press.
Bilgrami, A. (2012). The unique status of self-knowledge. In A. Coliva (Ed.), *The self and self-knowledge* (pp. 265–278). Oxford: Oxford University Press.
Coliva, A. (2016). *The varieties of self-knowledge*. Basingstoke: Palgrave Macmillan.
Gendler, T. (2008). Alief and Belief. *Journal of Philosophy, 105*, 634–663.
Gilles, A. (2001). A new solution to Moore's paradox. *Philosophical Studies, 105*, 237–250.
Green, M., & Williams, J. (2007). Introduction. In M. Green & J. Williams (Eds.), *Moore's paradox: New essays on belief, rationality, and the first person* (pp. 3–36). Oxford: Oxford University Press.
McGlynn, A. (2013). Believing things unknown. *Noûs, 47*, 385–407.
Sorensen, R. (1988). *Blindspots*. Oxford: Oxford University Press.
Williams, J. (1979). Moore's paradox: one or two? *Analysis, 39*, 141–142.
Williams, J. (1996). Moorean absurdities and the nature of assertion. *Australasian Journal of Philosophy, 74*, 135–149.
Williams, J. (1998). Wittgensteinian accounts of Moorean absurdity. *Philosophical Studies, 92*, 283–306.
Wright, C. (2013). A plurality of pluralisms. In N. J. L. L. Pedersen & C. Wright (Eds.), *Truth and Pluralism: Current Debates* (pp. 123–153). Oxford: Oxford University Press.

[6] If one favoured Coliva's verdict about the case over mine, would that spell the end of the minimalist account of the paradox I've sketched in this section? I think not, but I lack the space to take this issue on, and so I'll need to rest content with the stance taken in the text.

[7] I am grateful to Lukas Schwengerer and Kegan Shaw for discussion of Coliva's book, and to an audience at a workshop on self-knowledge at the University of Uppsala in April 2018, organised by Carl Montan, for feedback on this material.