Colivan Commitment vis-à-vis Moore's Paradox
T. Parent (Virginia Tech)

Annalisa Coliva's *The Varieties of Self-Knowledge* (2016) is the most thorough defense

available of *pluralism* about self-knowledge.[1] Pluralism, moreover, is correct. Therefore,

Coliva's is the most thorough defense available of the correct view of self-knowledge.[2]

But while this praise is sincere, it is my unfortunate task to quibble with some of her

details. And in this, I will mainly focus on Coliva's treatment of Moore's paradox in the

appendix. While this may seem restrictive, the issues addressed are in fact fundamental to

Coliva's overall position, concerning "commitments" (an important class of intentional states)

and self-knowledge thereof. The paper ends by noting points of convergence between Coliva's

constitutivism about commitments, and the constitutivism in Parent (2017).

1. Constraints on Moore-paradoxicality

Moore-paradoxical sentences are widely recognized as having either the *commissive* or

omissive form. Specifically, if 'Bip' represents in logical form the thought *I believe that* p,[3] then

these forms are:

---

[1] All references to Coliva shall be in relation to this book, unless otherwise specified.

[2] My own book does not elaborate a pluralist account, yet its aim was not to give a *general* account of self-knowledge. Rather, its focus was on the sub-type of self-knowledge needed for rational self-reflection. Though on p. 40 it explicitly endorses pluralism as a general account, and refers the reader to Coliva (2016) for details. I regret that Coliva's book was published just when mine was going to press, and I thus did not have a chance to discuss the important points of convergence. I attempt to remedy this in the final section.

[3] The underlined indexicals '$\underline{i}$' and '$\underline{I}$' express the "essential indexical" in the sense of Perry (1979); cf. Coliva on the third-person in connection with Moore's paradox, top of p. 248.

(COM) $p \wedge B\underline{i}\sim p$ -or- $\sim p \wedge B\underline{i}p$

(OM) $p \wedge \sim B\underline{i}p$

"Moore's paradox," in its original formulation, consists in the fact that (i) such forms are logically consistent, yet (ii) assertions with these forms naturally appear self-defeating in some obvious sense. Consider, e.g., my asserting of:

(G1) "There's no God, but I believe there is."

(G2) "There is a God, but I don't believe there is."

With respect to such assertions, the task is to explain (ii) while respecting (i).

One excellent feature of Coliva's discussion is the recognition that some instances of (COM) and (OM) are *not* self-defeating in the pertinent sense.[4] Coliva here offers the example of Jane, who explicitly believes that her partner Jim is faithful to her, even though Jane's behavior suggests that she unconsciously harbors doubt (e.g., surreptitiously checking his email, his phone, etc). Suppose that one day these doubts about Jim come into her awareness. Then, Jane might judge or assert:[5]

(J1) I believe that Jim is not faithful to me, even though he is.

Let us note also that Jane could have judged/asserted something with the omissive form as well:

(J2) Jim is faithful to me, although I don't believe that he is.

---

[4] Wittgenstein already observed that instances of (COM) and (OM) must be asserted, and asserted sincerely, before they become Moore-paradoxical. But we presuppose here that any utterances are indeed sincere assertions.

[5] Moore himself focused only on *assertions* having the form of (COM) or (OM). But Coliva (following Sorensen 1988) correctly sees that unspoken judgments can be Moore-paradoxical as well (thus revealing the limits of Moore's own explanation in terms of conversational norms). I shall thus treat propositions like (J1) as the possible content of an assertion or of a judgment.

But these are not equivalent propositions, and for now, let us follow Coliva in focusing on (J1).

The key observation is that (J1) has the logical form of (COM), yet it is quite appropriate for her to assert/judge it under the circumstances.[6] As Coliva notes, Jane's total state of mind would manifest some degree of irrationality, to be sure. Yet (J1) itself is not irrational: It is an entirely rational assessment, albeit regarding a somewhat irrational state of mind. At the least, (J1) is not "self-defeating" in the Moore-paradoxical sense.

One reason this is so, according to Coliva, is because the self-attributed belief about Jim's infidelity is merely a *dispositional* belief. Coliva here means to contrast "dispositional" intentional states, not necessarily with "occurrent" states (p. 28), but rather with her specially defined class of "commitments." In the first instance, however, dispositions are identified by the following distinguishing-marks. Such intentional states are (paraphrasing p. 28):

(a) not the result of a conscious deliberation, like a judgment, based on considering and assessing (or even being able to assess) evidence for the truth of *p* (or for the worth of *p*, etc.)
(b) not within one's direct control, being rather something one finds oneself "saddled" with;
(c) not something one will be held rationally responsible for (yet one is sometimes held "responsible" for them in a broader sense; see p. 32, n. 14).

However, Coliva gives some further, informative descriptors of "dispositional" states in her sense, which we can list as follows. These states are:

(d) often credited to a-conceptual animals to make sense of their behavior (p. 28),
(e) not always self-known (p. 29),
(f) often self-known only through a process of self-interpretation (e.g., on the therapist's couch, when explaining one's behavior) (p. 29),
(g) often predicable, either through inference or simulation, as having an influence under specific types of circumstance (even though they are not under one's direct control, as per (b)) (p. 30),

---

[6] In fact, Coliva is drawing attention to the corresponding *belief* in Jane, rather than any *assertion* on Jane's part. But I shall gloss this for now, and address only later the difference between belief and assertion.

(h) characterized in functionalist terms (p. 255, n. 28), i.e., they are explanatory mediators between sensory inputs and behavioral outputs (cf. p. 256),

(i) not intrinsically normative (p. 255, n. 28; p. 256). (This means they lack the relationship to norms that "commitments" have; see (f′) below for more.)

Perhaps there is some conceptual overlap between (a)-(c) and these further descriptors; regardless, the additions help provide further clarity.

Coliva's idea, then, is that (J1) fails to be Moore-paradoxical, partly because the self-attributed belief is just a dispositional belief. She writes "there is nothing paradoxical in finding out and, therefore, in self-ascribing a given belief, which contradicts what one explicitly judges to be the case, as long as that self-ascribed belief is, in fact, a mental disposition" (p. 255).

Naturally enough, then, Coliva adds that a Moore-paradoxical case must be concerned with a *commitment* (in her sense) rather than just a disposition. This follows immediately if dispositions and commitments exhaust the types of intentional states (although the two categories are not exclusive; see p. 37.) Regardless, let us here document what Coliva counts as a "commitment." Such intentional states, which also can include desires, intentions, etc., as well as beliefs, are (paraphrasing pp. 31-32):

(a′) the result of the subject's (mental) act of judging, based on a (possibly unconscious, possibly forgotten (p. 35)) consideration of the evidence, where the commitment stands as what the subject thinks she ought to believe (or ought to desire, or ought to intend, etc.), though of course she may be wrong about these "oughts" (see p. 36),

(b′) governed by norms of theoretical and practical reasoning (i.e., they are "normatively constrained"),

(c′) normatively constrained, even according to the subject's own point of view;

(d′) states for which the subject is rationally responsible (but see again p. 32, n. 14).

As before, Coliva's text offers further, informative markers for this type of state, beyond the official criteria. (Again, there may be conceptual overlap, but these still help in clarifying matters.) She says that commitments are:

(e′) in the subject's control (p. 31), although this does not mean one can adopt any commitment by an arbitrary act of will (p. 33),

(f′) co-instantiated with "dispositional elements," e.g., a belief-commitment comes with dispositions to use the belief as a premise in reasoning, to allow challenges to the belief, to produce evidence in response to challenges, yet to possibly withdraw the belief if the challenges are evidentially powerful enough, and to adjust other relevant commitments in light of the withdrawal (per a "local holism" of one's commitments) (p. 33),

(g′) had by some subjects who do not pursue the action recommended (if any) by its content (p. 259). E.g., It is possible for one to have a commitment to the ethical belief "I should donate to charity," even if one never actually donates to charity. (It is a case of "failing to live up to one's commitments.")

(h′) are first-personal rather than third personal in nature. Minimally, this means that if one has incompatible social (third-person) commitments, as when one makes conflicting promises to different parties, this does not suffice for incompatible first-personal commitments (p. 260).

So again, the point is that an assertion/judgment with the relevant logical form is Moore-paradoxical only if the self-attributed intentional state is a commitment.

Coliva has one further necessary condition on a Moore-paradoxical case. She says that something of the form (COM) or (OM) is Moore-paradoxical only if the self-ascribed state is *self-known* (p. 255). This is quite in order, for Coliva elsewhere says that an intentional state is a commitment (in her sense) only if it is self-known; this is necessary for the commitment to satisfy (f′) (p. 190). Thus it seems that the self-knowing requirement on Moore-paradoxicality is already built in, since we were restricted to commitments. Still, this reminds us that 'commitment' here is a quasi-technical term, for "commitments" in the ordinary sense are not always self-known. (My philosophical positions, for example, can imply "commitments" that I may be dimly aware of at best.)

Regardless, it now seems that the committing-feature of a self-ascribed state is ultimately what distinguishes the Moore-paradoxical assertions/judgments with the relevant logical forms. Again, such states are *ipso facto* self-known, and are not merely dispositional. However, Coliva is aware that further constraints on Moore-paradoxicality might be desired. In particular, she considers the question of "whether one can knowingly and willingly hold inconsistent

commitments" (p. 258). If such inconsistency is possible, then there may be other instances of

(COM) or (OM) that are not self-defeating in Moore's sense. On this score, I cannot help but

note that Coliva herself seems friendly to dialethiesm,[7] and such a person may find herself

saying:

> (D1) I believe the Liar is true, and it is not true.

The speaker here incurs "inconsistent commitments," but (D1) does not seem self-defeating in

the relevant sense, despite having the logical form of (COM). In a similar vein, I can imagine an

eliminative materialist asserting something like:

> (E1) There are no beliefs, and I don't [even] believe that.

No doubt, there is a paradoxical air here as well, yet the eliminativist does not seem to be

engaging in self-defeat (despite what some uncharitable critics may suggest[8]). As a third

example, a quantum physicist may find themselves saying in a non-self-defeating way:

> (QM) This electron has no definite location, and yet I can't believe that.

This would be the report of a certain cognitive limitation or "block," but it does not seem self-

defeating in the sense pertinent to Moore. Finally, one might imagine a devotee of the *Iśā*

Upaniṣad holding a commitment to the overtly contradictory verse 5:

> (IU) *Ātman* moves, yet I also believe that It moves not.

Perhaps I could persuaded that this is Moorean self-defeating, although it is unclear. At the least,

it would be atypical as a case of Moorean self-defeat, since "self-defeat" might be exactly the

---

[7] See Coliva & Moruzzi (forthcoming)

[8] See Parent (2017, ch. 9) on one strategy for how the eliminativist may avoid self-defeat. See also Joyce (2013) for a different strategy.

point: The speaker's aim may be to suggest that one *should not* attempt to speak intelligibly about *Ātman*.

Now in fact, Coliva argues that it is *impossible* to knowingly and willingly have inconsistent commitments (pp. 265-267). She considers here a man who makes conflicting promises, requiring him to be two places at once. Her analysis of this example seems quite reasonable, but it is one example. So I am curious what she might say about other cases like those above, and whether these require us to put further constraints on Moore-paradoxicality.

Again, Coliva's stated view comes to this: An assertion/judgment of the form (COM) or (OM) is Moore-paradoxical only if the self-attributed state is a commitment (in her sense) rather than a mere disposition. Note that the converse claim is not part of Coliva's view. Given an assertion/judgment of the form (COM) or (OM), it is not part of Coliva's view that Moore-paradoxicality is inevitable if the self-attributed state is a commitment. This is as it should be. For there seem to be cases where the self-attributed state is a commitment, yet there is no Moore-paradox, given that the *other* state is merely dispositional.

Thus, let us revise the omissive case where Jane sincerely asserts (J2). Suppose here that she is prompted to assert the first conjunct in a spontaneous, knee-jerk way, abetted by the fact that she strongly *wants* to believe in Jim's fidelity, is strongly *habituated* in believing in his fidelity, etc. But imagine now that she follows with the second conjunct "…although I don't [really] believe it", as a way of admitting a considered agnosticism on Jim's fidelity. (Assume the second conjunct is here not just an indirect or "coy" way of expressing positive disbelief.) Then, the second conjunct plausibly expresses some kind of Colivan commitment. Yet the case is not Moore-paradoxical because the *first* conjunct expresses a kind of dispositional belief.

As I say, all this is consistent with Coliva's account, although it may indicate a small lacuna in the constraints on Moore-paradoxicality. Perhaps Coliva really should say that a Moore paradox requires *both* conjuncts to indicate commitments rather than just dispositions. (This is, in fact, confirmed in her explanations of Moore-paradoxes; see section 2 below). However, our variant on the Jane example can also suggest that Colivan commitment is ultimately unnecessary to a Moore-paradoxical case, as I shall now explain.

Consider that it seems *possible* for an assertion of (J2) to be Moore-paradoxical. After all, the initial conjunct looks "inconsistent" with the agnosticism of the second conjunct. Now on Coliva's view, where (J2) is self-defeating, it requires "[t]aking the…doxastic conjunct…as the self-attribution of a commitment" (p. 56). One minor difficulty is how exactly to interpret the conjunct 'I don't believe he is [faithful]' as attributing a commitment. For it seems not to ascribe a belief as much as attribute the *absence* of a belief (viz., in Jim's fidelity). But so be it: Assume the commitment here is not a belief but rather some other intentional state. (This, however, is a noteworthy difference between Coliva's treatment of the omissive vs. commissive paradoxes.) Yet if the commitment in question is not a belief, what kind of intentional state is it?

The natural reply would be that it is a commitment to the agnostic-attitude, re: Jim's faithfulness. It would be a commitment to "neither-believing-nor-disbelieving" the proposition that Jim is faithful. (Coliva confirms this by speaking of "open-mindedness" in this connection; see, e.g., p. 264.) Note well: It is clear how such a commitment would create "self-defeat" if Jane asserts/judges (J2). The first conjunct would suggest a commitment that is literally inconsistent with it, for the first conjunct suggests Jane's commitment to positively *believing* the proposition. So on the surface, Coliva's account is faring just fine.

But the difficulty is that Jane's agnosticism might not fit the descriptions of "commitment" that Coliva has provided. We need to approach this point slowly however. So for starters, imagine first that we are not dealing with the Moore-paradoxical version of Jane's case, but rather, a case where Sam asserts plain-old-agnosticism about her partner's fidelity: "I'm not sure…I don't believe [but nor do I disbelieve] that Alex is faithful." Now, if there is a *commitment* to "neither-believing-nor-disbelieving," then according to (f'), it comes with a disposition to defend the attitude from challenges. But even a considered agnosticism about your partner's fidelity might not co-occur with a disposition to *defend* such agnosticism.

So for instance, suppose a friend challenges Sam's reflective agnosticism by highlighting some of Alex's behavior which evidences strong devotion to Sam. Sam might then *eagerly* relinquish her agnosticism in favor of belief, and thank her friend for emphasizing behavior that she had not properly accounted for. Surrendering her agnosticism in response to the challenge might even be *hasty*—and this would further the case that in no meaningful sense did Sam have a disposition to defend agnosticism.

Perhaps this would go to show that Sam's "reflective" agnosticism was not really a *commitment*, but more of a disposition. Perhaps, but this would concede the larger point I am trying to make. For my ultimate point is—if we graft Sam's type of agnosticism onto *Jane* when asserting (J2)—then the Jane case remains Moore-paradoxical even though the doxastic conjunct is not self-attributing a *committed* agnosticism. This is to say: Even if Jane has the kind of agnosticism that she is eager to relinquish (thus flouting (f')), it would still be Moore-paradoxical for her to sincerely assert/judge such agnosticism alongside the proposition that Jim is faithful.

Now in fact, the dialectics can get complicated here…for as we saw, Coliva holds that it is cognitively impossible for a Moore-paradoxical case to be real. Coliva's argument depends on

the premise that the commitment in the doxastic conjunct will "undo" the commitment in the other conjunct (p. 262). And so, her reply to my "new Jane" example may be that I am describing an impossibility, viz., a case where Jane is committed both to the belief that Jim is faithful, and to agnosticism on the matter. Or, insofar as I am describing a possible case, Coliva may protest that I am not describing a Moore-paradoxical case. It may just be a case where Jane feels some degree of uncertainty, while committing to belief in Jim's fidelity regardless. (Uncertainty here may only be a disposition, or perhaps a mere attendant feeling.)

However, part of my point is that Jane's agnosticism would not seem like a Colivan commitment, given that (f′) is flouted. And if not, then the case cannot be discounted as impossible on the grounds that one commitment is undone by a different commitment.

The other part of my point is that Jane's sincere asserting of (J2) would be Moore-paradoxical, despite her highly unstable (and entirely possible) agnosticism. So the problem Coliva faces is this. The new Jane example suggests that Colivan commitment in the doxastic-conjunct is inessential to Moore-paradoxicality. The case suggests that a Colivan commitment can be absent, and yet Moore-paradoxicality remains.[9]

---

[9] For the record, the "new Jane" example may not be the only example that creates the problem for Coliva, perhaps not even the most compelling example. Imagine, for instance, that Jane's assertion of (J1) is simply a confused, hurried, statement, which was not preceded by any weighing of evidence (even unconsciously). That is, suppose it is just Jane spontaneously blurting out something with the form (COM). The assertion still strikes me as Moore-paradoxical, even without condition (a′) on Colivan commitment being satisfied. We can also imagine Jane being disposed to immediately retract the assertion, thus making (f′) again unsatisfied. The assertion still is plausibly Moore-paradoxical, despite being the unreflective, flaky speech-act that it is. Jane's immediate retraction may show that she is being less irrational than if she stood by the assertion in the manner of a "commitment." But it would still

2. Explaining the self-defeat

So far, we have just been discussing the requirements on a Moorean assertion/judgment to exhibit Moore-paradoxicality. But we have yet to consider Coliva's *explanation* of the paradox, in those cases where the paradox indeed shows up. Now in fact, Coliva offers us two explanations, concerning somewhat different phenomena. Moore's original goal, recall, was to explain how assertions of the form (COM) or (OM) are self-defeating, even though they are logically consistent. And Coliva, following more recent discussions, has widened the *explanandum* to include *judgments* of the form (COM) or (OM). But what is meant exactly by "self-defeating?" Most writers have assumed that this means that the assertions/judgments are *obviously irrational* to a degree that approaches overt logical inconsistency (even though they are consistent, strictly speaking). And indeed, Coliva offers one explanation of Moore-paradoxicality which upholds this thought. But she offers a second explanation that goes beyond this. Above, we considered how Coliva sees Moore-paradoxicality as, in fact, cognitively impossible. She thus offers a second explanation for why this is so.

Both explanations, however, centrally feature the commitment-feature of the relevant intentional states. Thus, in considering a Moore-paradoxical version of Jane's assertion (J1), Coliva suggests that the commitments come with dispositions to reason from inconsistent premises. And the irrationality of the latter explains the irrationality in Moore-paradoxical assertions/judgments. She writes:

> Being committed to [P], one ought to use it as a premise of one's practical and theoretical reasoning. Hence, if Jane's judgment (or assertion) contained the self-ascription of a commitment…to hold that her husband is unfaithful…then Jane ought to use [My husband is unfaithful to me] as a premise of her reasoning. By also assenting to its

---

seem to have the basic Moore-paradoxical feature of "self-defeat," even if Jane subsequently flees the self-defeat after a moment's thought.

negation, however, she would commit herself to using (knowingly and willingly) *that* content as a premise of her reasoning. Thus she would commit herself (knowingly and willingly) to reasoning from contradictory premises. And *this* would be irrational. (p. 260)

So it is the short route between inconsistent commitments and defective reasoning which explain why Moore-paradoxical judgments/assertions are patently irrational, despite their logical consistency. (And notice that the explanation has both conjuncts of Moorean assertions/judgments indicating commitments, as per my earlier suggestion.)

Coliva's point here strikes me as reasonable, despite my earlier worry about the necessity of Colivan commitment in Moore-paradoxicality. (My honest opinion is that she may be right about most cases regardless.) Yet even if we accept her explanation of the irrationality, the explanation might ultimately be multi-factorial. Consider, for instance, that the two commitments not only would encourage contradictory premises in reasoning, but also contradictory testimony (even in the absence of reasoning). Thus, assume Jane's Moorean judgment/assertion reflected inconsistent commitments. Then, if she were asked "Is your husband faithful to you?," her commitments would also support the response "yes and no" (interpreted as strictly contradictory). Since this testimony is obviously informationally useless, it is irrational to attempt communicating information by such means. Thus, one further reason for the strong sense of irrationality in Moore-paradoxical cases could owe to its connection with silly attempts at communication.

Regardless, this too would feature Colivan commitment as central to explaining the irrationality connected with Moore's paradox. But as noted, Coliva is also concerned to explain the alleged cognitive impossibility of Moore-paradoxical cases, based on the alleged cognitive impossibility of inconsistent Colivan commitments. Now at first, the two *explananda* may seem to sit uncomfortably together. After all, if Moore-paradoxical cases are impossible, then why

worry about Jane's reasoning from contradictory premises? But charitably interpreted, Coliva's point about contradictory premises was to explain why Moore-paradoxical cases *would* be irrational, if they *were* genuinely possible. That is quite consistent with saying, in addition, that they are not genuinely possible.

As for the explanation of the alleged impossibility, Coliva's thought is that if Jane asserts/judges (J1) as an attempt to have inconsistent commitments, a Colivan commitment to Jim's infidelity would "undo" the first Colivan commitment to his fidelity. So any such attempt will fail. On the general phenomenon, Coliva writes:

> [a] commitment could so much as exist only as long as no known and willing assent to the negation of its content were in place. Hence, if the doxastic conjunct is a self-ascription of a genuine commitment, no assent to its negation is possible. Conversely, if such an assent is in place, then there cannot be any commitment to [P] in the first place. (p. 264)

(I assume in this passage that "assenting" to a content is the same as committing to a belief with that content.)

Take heed: It would be quite unfair to object that people often have inconsistent commitments, hence, such commitments are possible. As Coliva elaborates (pp. 265-267), those are plausibly cases where the person really has forgotten one commitment, or relinquished a commitment, or (mis/)understood the commitments as consistent. And in a Moore-paradoxical case, such things are not *bona fide* possibilities (at least for cognitively typical agents), since the inconsistent commitments are invoked "in the same breath," as it were.

I myself am inclined to concede all this, mostly because Colivan commitments are not necessarily commitments in the ordinary sense, as already noted. And yet there may well be a type of intentional state that counts as a "commitment" defined in Coliva's way, which can be undone once an inconsistent commitment arrives on the scene. It may not be clear, however,

what justifies believing in such an "undoing" feature. Nothing in the descriptors (a′)-(h′) clearly imply it. Is the "undoing" capacity meant to be supported by an inference to the best explanation? If so, it may be a problem that some philosophers would reject the *explanandum*. After all, if one doubts the cognitive impossibility, then the "undoing" capacity is not needed to explain such an impossibility. Granted, there may be a conversational norm of interpreting a speaker's commitment as "undone," once an inconsistent commitment is assented to. But I am unsure why Coliva believes that this might reflect a deeper cognitive reality, where one intentional state causes another of the same type to go out of existence.

3. Convergence

I have been paying Coliva the philosopher's awkward compliment of attending to her views enough to criticize them. But let me also compliment her more directly, by noting the many striking points of agreement between her work and mine. Not only are we both avowed pluralists about self-knowledge, we also agree on *constitutivism* about self-knowledge for those non-dispositional intentional states employed in reflective reasoning. Coliva's characterization of constitutivism is as follows (paraphrasing p. 164):

> (Con) In conditions C, one believes (/desires/intends) that *p* (/to ϕ) iff one believes that
>
> one believes (/desires/intends) that *p* (/to ϕ).

The claim is that, in the right conditions, a true *higher-order belief* about one's belief/desire/intention is necessary and sufficient for that belief/desire/intention. The relevant conditions here (besides obvious things like sincerity, conceptual competence, etc.) are where the first-order states are involved in our "deliberating what to believe, desire, intend to do, and so on" (p. 165). Now we would both elaborate further on the requisite conditions C in the

constitutivism thesis. And, though my (2017, ch. 7) book focused more on the right-to-left direction of (Con), the arguments described conditions where there was a *co-occurrence* of a first-order state and its correct self-attribution. Moreover, the joint occurrence of the states trivially implies the material biconditional regarding their occurrence. So although it may not be entirely obvious, the book is squarely in the constitutivist camp as Coliva has defined it.

We are not entirely agreed in some of the details, naturally, but there is remarkable agreement in some of the details as well. For instance, once the right conditions are specified, Coliva would agree that the right-to-left direction of (Con) holds of *necessity*, (see, e.g., p. 200). From my point of view, this qualifies her as a type of *infallibilist* about the relevant set of self-attributions. But granted, Coliva may be hesitant to join me in applying that much maligned label to her view (understandably so…although I wouldn't mind the company).

Another point of convergence is that we explain the constitutive thesis in fundamentally the same way. For one, we both deny that there is an "epistemic achievement" involved. That is, we each say that the "match" between the first-order state and the second-order attribution (in the right conditions) is not due to any epistemic act of weighing evidence, making inferences, testing hypotheses, etc. But more strikingly, Coliva and I both employ an "anti-intellectualist" view of linguistic knowledge, in order to elaborate on this. According to Coliva, the truth of (Con) ultimately owes to a (conceptually competent) subject being *trained* by the community to replace her assertions of "*p*" with assertions of "I believe that *p*," and vice-versa. Such training does not presuppose self-knowledge of belief, since the the training is done "blindly," as she puts it. I take this to mean that such training is merely the reinforcement of certain dispositions for linguistic usage, based on a "stimulus-response" picture of linguistic conditioning. (I elaborate such a model in ch. 8, §7, of Parent 2017, by appeal to neural net models of language acquisition.)

Finally, Coliva and I concur that (Con) describes *synchronic* self-knowledge only, i.e., knowledge indexed to some specific time. And there is no assumption that such knowledge at one time will carry over to another time. But since diachronic self-knowledge is necessary, e.g., in the temporally extended process of deliberation, we are both pushed to go beyond (Con), strictly understood. This is where Coliva emphasizes how a subject herself can deploy "third personal" methods (e.g., observation and abduction from one's linguistic and non-linguistic behaviors), which of course accords with her general pluralism about self-knowing. I, too, resort to a kind of "extrospection" based on one's own linguistic behavior, although (Con) still does some work in explaining why the behavior is evidentially relevant. But in the case of diachronic self-knowing, we are unified in the view that first-person constitutive authority does not exist, strictly speaking—even though we are equally unified in saying that such authority is a real and robust synchronic phenomenon.

References

Coliva, A. (2016). *The Varieties of Self-Knowledge*. Palgrave.

Coliva, A. & Moruzzi, S. (forthcoming). Dialetheism in action: A new strategy for solving the equal validity paradox? In A. Rieger & G. Young (Eds.) *Dialetheism*. Springer. Draft available at https://www.academia.edu/34861271/Dialetheism_in_action_a_new_strategy_for_solving_the_equal_validity_paradox_Draft_1.3.

Joyce, R. (2013). Psychological fictionalism. *The Monist* 96: 517-538.

Parent, T. (2017). *Self-Reflection for the Opaque Mind*. London: Routledge.

Perry, J. (1979). The problem of the essential indexical. *Noûs*, 13, 3–21.

Sorensen, R. (1988). *Blindspots*. Oxford UP.