Just Another Article on Moore's Paradox, but We Don't Believe That

Abstract: We present counterexamples to the widespread assumption that Moorean sentences cannot be rationally asserted. We then explain why Moorean assertions of the sort we discuss do not incur the irrationality charge, appealing to the dual-process theory of the mind and a contrast between the conditions for ascribing beliefs to oneself and the conditions for making assertions about independently existing states of affairs. We conclude by contrasting beliefs of the sort we discuss with the structurally similar but rationally impermissible beliefs of certain psychiatric patients.

Keywords: Moore's paradox; Dual-process theory; first-person evidence; third-person evidence; irrational beliefs; a-rationality

A Moorean sentence is a sentence such as, "*P*, but I don't believe that *p*," or "*P*, but I believe that *not-p*," commonly exemplified by, "It is raining, but I do not believe it is," or, "It is raining, but I believe it is not raining." Moorean sentences are thought paradoxical, because allegedly they cannot be either asserted or thought without absurdity, yet both conjuncts of a Moorean sentence can be simultaneously true: it can be true that it is raining and that I don't believe it is, and it can likewise be true that it is raining but that I believe it is not.[1] Other students of Moore's Paradox see their main task as that of explaining just why a Moorean sentence cannot be rationally asserted.[2,3] But is it true that asserting a Moorean sentence (or thinking the corresponding thought) is always irrational?[4] Our purpose in this brief communication is to challenge that assumption.

Before we turn to our task, we note that others have given examples—very different from the type of example we will present—of Moorean sentences that allegedly can be asserted without

[1] Moore 1944, pp. 203-4. While Moore was, to our knowledge, the first philosopher to draw attention to these sentences, the label "Moore's Paradox" is attributed to Wittgenstein.

[2] Moore himself (1944, pp. 203-4) argued that since people in general do not assert *p* unless they believe *p*, in asserting *p*, we *imply* that we believe *p*, despite the fact that in asserting *p*, we neither assert that we believe *p* nor is our believing *p* entailed by our asserting *p*. Contemporary solutions tend to center on the relationship between first-order beliefs and second-order awareness of first-order beliefs (see Rosenthal 1986; Shoemaker 1995). The relationship between first- and second-order beliefs in connection to Moore's paradox is discussed in Larkin 1999; Kriegel 2004; Kind 2003; Kobes 1995; Fernández 2005; Lee 2001; and Williams 2006, among others.

[3] Heal (1994), in a discussion of Wittgenstein's proposal, suggests that the problem is to explain why Moorean sentences cannot be uttered with sincerity or thought with conviction *at all*, and not simply why they cannot be asserted without irrationality. We note, therefore, that if we succeed in convincing the reader that Moorean assertions can be made without irrationality, the weaker claim that they can be made at all will follow.

[4] For an unusual example, see Sorenson's (2001) *Vagueness and Contradiction*, pp. 28-9, where he argues for the irrationality of Moorean sentences even with respect to borderline cases.

absurdity.[5] Whether any of those examples are Moorean is a matter of dispute.[6] Whatever the verdict, the sorts of cases we have in mind are different and of independent interest. The phenomena we wish to discuss, involving first- vs. third-person evidence for belief ascription, and the dual-processes theory of mind, occur naturally and frequently. Thus, making sense of these phenomena is crucial to understanding what makes a belief rational or irrational, or so we wish to argue.[7]

I.

IA.

The Grand Canyon Skywalk—a glass, horseshoe-shaped structure complete with translucent walls—extends 70 feet from the rim of the Canyon into space. Luke, although attracted to the promise of spectacular views and assured of the structure's stability both by its engineers' credentials and by witnessing many safe crossings of fellow humans, experiences—much like other tourists—extreme fear and reluctance to cross. But, again, Luke has objective empirical evidence that the Skywalk is safe.[8] We can represent Luke's experience with a Moorean sentence: "It is safe, but I don't believe it is."

We wish to suggest that this Moorean sentence can be asserted without absurdity. Borrowing a distinction from Richard Moran,[9] we can say that the fearful Luke has a theoretical, third-person-type endorsement of the proposition that the structure is safe, while from the first-person point of view, he has a compelling *seeming,* giving rise to the belief that the Skywalk may not be safe or to doubt that it is safe, which is why he experiences fear and reluctance to cross. The first-person seeming may not give rise to an actual belief in the negation of the proposition based on third-person evidence, i.e., Luke need not believe that the Skywalk is *not* safe, as per what is known as the "commissive" version of Moore's paradox. But neither is it the case that he believes it *is* safe simpliciter. If his only belief were that it *is* safe, he would not show reluctance to cross. Taking all of this into account, we can present Luke's state of mind most fully as follows: The structure is safe (my evidence suggests that it is; I'm allowing my children to cross), but I don't believe it is safe (I can't embark on the crossing).

But is it true that the Moorean conjunction, "It is safe, but I do not believe it is," can be held without irrationality? It depends on what one means by "rational" and "rationality," but on an ordinary understanding of these notions, the answer is that it can. Ordinary intuition suggests that Luke is not

---

[5] Crimmins 1992; Turri 2010; Pruss 2012. Crimmins' example involves learning that a person you know to be very intelligent is also someone whom, under a different guise, you consider an idiot. You can then assert sincerely, "I falsely believe you are an idiot," which is arguably equivalent to the Moore-paradoxical, "You are not an idiot, but I believe you are an idiot." Turri gives an example of an eliminativist about belief who nonetheless cannot abstain from having beliefs. Thus, she may say, "It's raining, but I don't believe it is" (since there are no beliefs). Pruss's main example involves a therapeutic case: an expert analyst persuades me that *p* is the case, but that I don't believe that *p*. Pruss does not fill in the details, here, but we could imagine, for instance, that the analyst persuades me that although my parent is not to blame for X, I believe the parent is to blame for X.

[6] Rosenthal (2002) and Stoljar & Hajék (2001) take issue, for different reasons, with Crimmins's case. Pruss 2012 is skeptical of Turri's example.

[7] There is a body of literature on cases involving tension between explicit judgment and behavioral and emotional attitudes (for instance, the explicit judgment that all races are equal coupled with an implicit racist bias with behavioral manifestations). These cases have been labeled "dissonance" cases. Discussion of them has proceeded largely independently of the debate on Moore's Paradox, with the exception of Gertler 2011 and Borgoni 2015, who interpret (some) instances of dissonance as Moore-paradoxical cases. The sorts of examples that interest us are a subset of dissonance cases. We will have something to say about Gertler's and Borgoni's discussions later (see note 14 below).

[8] An example along these lines was previously discussed in Gendler 2008.

[9] Moran 2001.

irrational. We do not wish to simply rely on intuitions here, however. Rather, our claim is that a careful study of Luke's belief-forming processes will illustrate that Luke's belief is not irrational. Here is why: behind the conflict between two types of evidence available to Luke—first- and third-person—there is a conflict between two cognitive systems that often function independently of each other, each furnishing different sorts of evidence. The first-person seeming results from "System I" processes, while the third-person assessment is an outcome of "System II" processes. Made popular recently by Daniel Kahneman,[10] *dual process* accounts of our cognitive operations can be found in earlier writings by William James, Freud, and several others.[11] System I processes are associational, non-reflective, and automatic. In themselves, these processes are neither rational nor irrational—rather, they are a-rational.[12] The question is whether it is rationally permissible for Luke to form a belief on the basis of a seeming resulting from these a-rational processes.

Our claim is that Luke's Moorean belief is rationally permissible. Even on a fairly demanding conception of rationality, when the evidence available to System I is extremely compelling, a person cannot be said to be irrational for not forming a contrasting belief on the strength of System II evidence only. And the System I evidence available to Luke *is* very compelling. Luke has precisely the attitude one would expect of a normal, rational person. It is natural for the fear one may experience in this situation to be so strong as to make it impossible for the person to only believe there is no danger. Of course, one could imagine a hyper-rational person whose first-person seemings are *always* in line with his or her third-person evidence, but there is no rational *requirement* to be *hyper*-rational. (Indeed, if someone experienced *no* fear in these circumstances, the likely explanation would be not hyper-rationality, but rather, a neural abnormality resulting in excessively low fear.)

If you have the lingering suspicion that Luke's fear and reluctance to cross despite all evidence that the structure is safe are in fact irrational, consider a more striking example. In *The Brothers Karamazov*, a schoolboy named Kolya Krasotkin, intent on impressing his friends, lies on the train tracks one day and waits for an approaching train to pass him over. Krasotkin is not on a suicide mission: he has done all the proper measurements and knows that the train won't touch him. Nonetheless, he faints while under the train (though he subsequently convinces his friends he felt no fear).[13] Again, we can capture Krasotkin's experience with a Moorean sentence: "The train can't hurt me, but I don't believe that." But there is no absurdity or irrationality about this thought and the corresponding assertion. Even if our opponent insists that Luke in the Skywalk case is simply being irrational in his reluctance to cross, and that he should fully base his belief on third-person evidence of the safety of the structure, surely it would be unreasonable to expect a person in Kolya's situation—a normal person, that is, rational but not hyper-rational—to be *so* responsive to third-person evidence

---

[10] Kahneman 2011. Kahneman summarizes his decades-long empirical research on areas of human thought and action shown to be "not-rational."

[11] James (1890) spoke of "associative" reasoning and reasoning-proper, or true reasoning. For Freud (1900), the "primary processes"—association-based and impulsive, very different from reflective, rationally-based secondary process assessments—predominate not only in psychiatric symptoms, but also in the normal mental life of children, as well as in much of the non-conscious and implicit cognitive activity of adults. See also Piaget 1926; Vygotsky 1987; Niesser 1963; Johnson-Laird 1983; Stanovich & West 2000.

[12] For a fuller account of the a-rationality and the primary processes, including their central importance in non-human animal cognition, see Brakel & Shevrin 2003; Brakel 2009.

[13] Dostoyevsky 2002, p. 501.

as to extinguish any residual idea that the train might, after all, hurt him or her, no matter how strong the evidence for lack of danger may be.

We wish to note here that other authors have made a different and relatively uncontroversial point regarding the possibility of a rational Moorean assertion: others have noted, correctly we think, that if a subject *were* to find himself or herself in a Moorean state, then the rational thing for the subject to do would be to acknowledge the tension in his or her own belief system. A subject on this view is irrational in being in a Moorean state but rational in recognizing that he or she is in such a state.[14] If that's all the rationality retained by the subjects, it does not amount to much. For a rational person who discovers a contradiction in his or her belief system is expected to change one or both beliefs. If both continue to persist, then in the usual case, a person is quite irrational.[15] Our claim is not that Luke and Kolya are rational in recognizing their own irrationality, but rather that they are not irrational in asserting *p* while failing to form the belief that *p* to begin with.

IB.

There are four alternative interpretations of the cases we have presented that we wish to consider, each of which implies that the cases may not be Moore-paradoxical. First, it can be claimed that the first conjuncts in Luke's and Krasotkin's utterances are not genuine assertions. Second, it can be argued that the second conjuncts are not reports of either genuine beliefs (as per the commissive version) or of genuine lacks of belief (as per the omissive version). Third, it can be maintained that the "I" implicit in the first conjunct is not the same as that which figures in the second conjunct. Fourth, a case can be made that the conjunctions uttered or thought by Luke and Kolya are instances of "in-between" beliefs. We shall take these points in order.

According to the first possibility, Luke's state of mind can be more accurately expressed in the following way: "The objective evidence implies that the structure is safe, but I just cannot believe that the structure is safe."[16] If so, then perhaps Luke's assertion—irrational or not—is not Moorean since

[14] Borgoni (2015, p. 108) writes, "This paper's view is that the person is irrational in being dissonant, although not irrational in asserting (or believing) a Moorean proposition." Gertler (2011, p. 140), similarly, writes, "Most importantly, Nick [subject in a dissonant state]'s willingness to take the psychologically difficult step of confronting the disparity between his belief and his reasons may reflect an especially strong *commitment* to norms of reasoning. So while one who endorses a Moore paradoxical thought is not ideally rational, the act of endorsing that thought may itself be one for which the thinker deserves cognitive *credit* rather than blame." Chislenko (2016, p. 687) also, in the context of a discussion of an anorexic who realizes his belief that he needs to lose weight is groundless, says, "His belief that he does need to lose weight may well be irrational; but the belief that he has that belief, and that it is false or that he should not have it, can itself be a rational one. The 'anorexic' belief may be the product of insecurity and a warped body image; but the Moorean and belief-akratic-paradoxical beliefs themselves can indicate an impressive and hard-won self-awareness." We note here that we accept Chislenko's interpretation of the specific example of the anorexic, because on our view, the anorexic's Moorean belief, unlike those of Luke and Kolya, is irrationally held unless the anorexic is sufficiently moved by the evidence in order to act on the evidence rather than the recalcitrant belief he is fat (see Section III of this paper). However, for Chislenko, it is true more generally (and not just of anorexics) that when it comes to Moorean states, rationality can accompany only the sort of second-order recognition of one's contradictory beliefs, not the subject's arrival at those beliefs or the beliefs' persistence in light of conscious recognition.

[15] Indeed, in some sense, one is more irrational in maintaining two contradictory beliefs knowingly than unknowingly, so what is gained in terms of second-order rationality may come at the expense of rationality at the first-order level, so that a knowingly irrational subject comes out more *self-aware* but not necessarily more *rational* overall compared to an unknowingly irrational one.

[16] We thank an anonymous reviewer for pointing out this option.

asserting that the objective evidence implies *p* is not the same as asserting that *p*. There is conceptual space for the following possibility: the agent has an akratic belief which is not Moore-paradoxical.[17] The agent with an akratic belief may fail to be sufficiently moved by the objective evidence and might maintain a belief contrary to the evidence. We have three things to say in response. First, one can argue that the revised statement *is* Moore-paradoxical, even if the first conjunct is a statement about what the evidence implies rather than a statement about what the world is like. Consider, "The objective evidence implies that it is raining, but I don't believe that it is." This sounds Moore-paradoxical to us, and we are not the alone here.[18] Second, the statement so revised appears to manifest irrationality on the part of a believer (since a rational believer, presumably, would be appropriately moved by what he or she takes the objective evidence to imply), and so showing that a believer in such a state is not irrational, as per our argument, would be a significant result. Third and most importantly, we wish to argue that there is no good reason to deny Luke assent to the proposition, "The structure is safe." Some aspects of Luke's behavior suggest that he assents to the proposition that the structure is safe, for instance, he is not trying to stop other people from crossing and allows his own children to cross, and he would presumably stop others (especially his own children) if he didn't assent to the proposition in question.

One could grant that the first conjunct is a genuine assertion and instead question whether the second expresses a genuine lack of belief. This is where we come to the second alternative interpretation above. One can argue that Luke's state of mind can be best expressed as follows, "The structure is safe, but I don't believe* that it is," where belief* is some state weaker than belief, such as alief[19] or "belief in a bodily way."[20] What about this suggestion?

Everything turns on what "belief*" is. Clearly, there are mental states whose contents may be in tension with our explicit assertions without the tension's generating a Moorean paradox. Thus, if I said, "Quantum entanglement is real, but I intuit that it is not," there would, arguably, be no paradox: our intuitions can go against what we assert or take to be the case on the basis of objective evidence without prompting us to form *beliefs* contrary to that evidence or else underwriting *lack of belief* in the assertions based on the evidence. Similarly, if I said, "I tremble because the Green Slime in the movie I am watching seems to be coming toward me, but no Green Slime is really coming toward me," there would be no paradox either: emotions such as fear can go against our judgments regarding what state of affairs obtains and what there is a reason to feel.[21] However, it is not the case that Luke simply finds it counterintuitive that the structure is not safe, nor that he finds the structure's alleged safety

[17] Chislenko (2016) argues that akratic beliefs need not be Moore-paradoxical, and that philosophers have been denying the possibility of akratic belief on the ground that accepting such a possibility commits us to accepting the possibility of Moore-paradoxical beliefs.

[18] For instance, Michael Huemer (2007, p. 146) suggests that the following is an instance of a Moore-paradoxical statement: "It is raining, but I have no justification for thinking so." Cf. Gallois 2007, pp. 166–7; de Almeida 2007, p. 56; Adler and Armour-Garb 2007, pp. 161–2).

[19] We thank an anonymous reviewer for this objection. The reviewer put the point in terms of "alief" rather than "belief*," but we have opted for a more general term meant to capture all relevant belief-like states, whether or not those meet the precise criteria of "alief" specified by philosophers.

[20] Jane Heal (1994, p. 15) entertains this possibility.

[21] On strong versions of cognitivism about emotion, emotions are judgments regarding what there is a reason to feel (Solomon 1980; Nussbaum 2001). Thus, to feel fear is to judge that there is something to be afraid of. In this sense, a conflict between an emotion and an explicit judgment can be interpreted as a conflict between two judgments. But we do not wish to rely on this view of emotion.

counterintuitive in addition to experiencing fear. Luke is reluctant or outright refuses to cross, which suggests that he is *moved* by the seeming that the structure isn't safe: the seeming either causes in him a belief that the structure is not safe or casts doubt on the belief the structure *is* safe.[22] This scenario can be contrasted with that of the moviegoer. It may seem to me that the Green Slime in the movie I am watching is coming toward me, but I have no impulse to leave my seat in order to protect myself. This suggests that the seeming does not move me sufficiently to generate a belief. Not so for Luke: it does not simply seem to him the structure is unsafe—he is extremely reluctant to act on the supposition that the structure is safe.

We've considered the possibility that the first conjunct in the sentences we discussed is not an assertion or that the second is not a belief report. There are two more alternative interpretations of the cases we have presented. One involves duality in the subject and can be found in Wittgenstein's (somewhat cryptic) remarks on Moore's paradox. Wittgenstein, as we interpret him, suggests first that in the usual case we cannot distance ourselves from our beliefs: where my beliefs go, I go, so to speak (since my beliefs do not exist independently of me), and I cannot distrust my beliefs and say such things as, "I falsely believe *p*." Neither can I distance myself from my beliefs in a way that allows me to infer my beliefs from my own behavior, as I might infer other people's beliefs from their behavior. However, having said this, Wittgenstein goes on to suggest that circumstances are imaginable in which I might achieve a distance from my own beliefs sufficient to make it possible for me to utter a seemingly Moore-paradoxical statement. Wittgenstein writes:

> If I listened to the words of my mouth, I might say that someone else was speaking out of my mouth.
>
> "Judging from what I say, *this* is what I believe." Now, it is possible to think out circumstances in which these words would make sense.
>
> And then it would also be possible for someone to say "It is raining and I don't believe it", or "It seems to me that my ego believes this, but it isn't true." One would have to fill out the picture with behaviour indicating that two people were speaking through my mouth.[23]

This points to the following possible reading of our cases: the "I" implicit in the first conjunct is not the same as the one that figures in the second conjunct. As Jane Heal notes in her discussion of Wittgenstein, the paradoxicality seems to vanish on this reading.[24] What of this possibility?

We do not know exactly what kind of case Wittgenstein has in mind (he does not tell us), but we wish to argue that the sorts of cases we have presented are not naturally described as involving two different subjects—me and my ego, or me and my body, or something like that. The "dual subject" interpretation would naturally fit a very different kind of case: that of Alien Hand Syndrome[25] or, perhaps, a case of bodily behavior caused directly by stimulating the brain. In these latter cases,

---

[22] "This structure is safe, but I doubt that," is a Moore-paradoxical utterance, as can be easily seen if we consider the parallel, "It is raining, but I doubt that it is."

[23] Wittgenstein 1953, p. 192.

[24] Heal (1994, p. 15) writes, "It is clearly presupposed in setting up the paradox that the 'I' spoken of in the explicit self description is the same as the person whose belief is expressed in the utterance as a whole…"

[25] Banks et al. (1989, p. 456) describe a patient whose "left hand would tenaciously grope for and grasp any nearby object, pick and pull at her clothes, and even grasp her throat during sleep." They go on to say that the patient, "slept with the arm tied to prevent nocturnal misbehavior. She never denied that her left arm and hand belonged to her, although she did refer to her limb as though it were an autonomous entity."

subjects don't identify with the behavior of their hands or bodies and speak of those behaviors as something not to be ascribed to *them*. Not so with Luke. It is not the case that Luke's *ego* fails to form the belief that the structure is safe where the ego is something distinct from Luke. It is *Luke* who refuses to cross, not just his ego.

There is a fourth and final option we wish to consider: it can be argued that these are cases in which the subject neither believes nor disbelieves either of the two conjuncts but rather is in an "in-between" state. Schwitzgebel, in a discussion of cases in which explicit judgments are in tension with (at least some key aspects of) behavior, argues that subjects in such instances are "in-between" two possible beliefs.[26] Schwitzgebel is not concerned with Moorean cases specifically but with a larger class of cases involving such tensions. A subset of the dissonance cases would resemble those we have focused on here in relevant ways. For instance, one of Schwitzgebel's examples involves a person allegedly persuaded by Stoic arguments to the effect that death is not bad, but who behaves in all other ways just like people who believe death is bad. Suppose this person did become aware of his or her fear-based behavior in the face of death. The "trembling Stoic," as per Schwitzgebel's label, can then presumably say, "Death is not bad, but I seem not to believe this." According to Schwitzgebel's interpretation, this subject neither believes nor disbelieves that death is bad. The idea is this: belief is a complex dispositional state: to believe is to be disposed to make certain kinds of explicit avowals as well as to behave in certain ways, whether deliberately, automatically, or by habit. Sometimes, a person, has some of the dispositions constitutive of a belief that *p* but lacks others. For instance, the fearful Stoic is disposed to sincerely avow that death is not bad, but he is also disposed to try to avoid death and to grieve the loss of a good person (and not only for his own sake but for the deceased's sake).

Should we say that the subjects in our cases are in an in-between state? We believe that the answer is "no." The "in-between" interpretation fails to capture the *conflict* between the two conjuncts of the Moorean utterance. The interpretation thus obscures the difference between Moorean cases and perfectly ordinary cases of uncertainty. Perhaps you think that your candidate is highly likely to win, but you are not entirely certain. In some ways, you behave as though your candidate will win: you cheerfully declare that your candidate will win, and perhaps, you place a bet. In other ways, however, you behave as though the candidate may not win: maybe, you move your businesses offshore before the election, because the other candidate has vowed to raise the taxes on locally operated businesses. You don't think it is very likely that the other candidate will win, but you think it may happen, and you want to cover all your bases. You are, thus, in an "in-between" state. But this is not a state of conflict in the relevant sense. You are not going to say, "My candidate is going to win, but I don't believe that." Rather, you will say, "I hope my candidate will win, and I think s/he is likely to win, but I am not certain." Since "in-between" states involve no conflict, they are, by default, not irrational. Moorean states, by contrast, *may* be compatible with rationality but they are not rational by default. A special argument is needed to demonstrate that a given Moorean state is compatible with rationality. We conclude from here that an interpretation of the cases described that collapses the difference between those cases and ordinary instances of uncertainty is unsatisfactory.

[26] See Schwitzgebel 2010.

IC.

We would like now to consider a different type of Moore-paradoxical case, involving classical conditioning, particularly aversive conditioning, in which negative stimuli are used to potentiate aversive reactions.

To demonstrate this second sort of case, we begin by offering the following illustrative example, presented in *Earth Magazine*: "Would you drink water out of a toilet, even if the toilet was never [and had never been] used?" [27] The question accompanies a photo of a man drinking toilet water from a toilet bowl. The article explains that this is the question asked in a museum display featuring brand new toilets "at the Exploratorium in San Francisco," noting that the issue is important because it's related to devising workable "wastewater recycling plans." The article reports that many museum visitors describe a feeling of disgust and an inhibition in drinking, while others refuse to drink. We wish to suggest that at least those museum goers who refused to drink, if not those that merely felt disgust, were in a Moore-paradoxical state that could be expressed with the sentence: "This water is perfectly clean, but I don't believe that it is."

Unlike the Skywalk and Train examples, the visual perceptual cues here indicate safety, not danger. One can see the cleanness of the toilet bowl drinking vessel in its gleaming, sparkling, newness. But surely for any adult residing in the developed world, toilets are regarded as the proper receptacle for excrement, and to a lesser degree menstrual blood, and vomit, never drinking water. Further, the thought of ingesting excrement (or the other mentioned bodily fluids) provokes an automatic response of disgust. Owing to that automatic response, toilets, like other excrement-associated elements, have been classically aversively conditioned to evoke disgust. And aversive conditioning, a widespread phenomenon across human and non-human animals alike, is a cognitive response predicated on System I primary process cognitive principles, featuring here associatively-based, a-rational similarity categorizations made spontaneously, without reflection.[28]

With this background, we can now revisit our central question: Can the above Moorean sentence ("This water is perfectly clean, but I don't believe that it is") be asserted without absurdity or irrationality? Our answer is again "yes." On the basis of third-person evidence, the speaker concludes that objectively speaking, the water is clean, and he/she believes this. But the speaker also acknowledges a belief that the water is contaminated, a belief which arises automatically and associationally, as a result of aversive conditioning.

II.

There are additional constraints on the rational assertibility of Moorean sentences. Rationality requires us to be sensitive to the kinds of evidence that ground beliefs, e.g., to be aware of whether the evidence is a first-person seeming or third-person evidence. This constraint underlies a norm governing assertion, one which requires us, in the event of a conflict between first-person seemings and third-person evidence, to base our assertions about the way the world is on third-person evidence. Self-ascriptions of belief, on the other hand, do not have to be exclusively based on that kind of evidence.

---

[27] *Earth Magazine*, January 28, 2013. Accessed November 28, 2016. http://www.earthmagazine.org/article/drinking-toilet-water-science-and-psychology-wastewater-recycling.

[28] As proof of concept, we would predict that young children who have not yet associated toilets with excrement would have no trouble drinking from the new toilet bowls (absent disapproving parents). And, adding to our argument, dogs find nothing at all aversive about toilet water, a fact to which most dog owners will attest. Some dogs actually prefer drinking from toilets. In these situations, dog training includes aversively conditioning the toilet. (For example, a drop of a liquid, bitter, nauseating, and repulsive to dogs, can be placed on the toilet rim.)

This is why Luke, Kolya Krasotkin, and the museum goer asked to drink water from a brand new toilet, while they can assert *a* Moorean sentence without absurdity, can only assert a *particular* sentence. They have to say, "It is clean/safe, but I don't believe it is." They *cannot* say, "It is not clean/safe, but I believe it is." The reason for this is that assertions about the way the world is must be based on sharable, public reasons. Our own beliefs, on the other hand, can be rationally permissible without being based exclusively on such reasons. To clarify this point further, suppose I meet our new neighbor, Francis, for the first time, and something about his smile and facial expression makes me feel mistrustful. You ask me what kind of person he is. I must say something like, "He was perfectly cordial, but I don't trust him." I cannot assert, "He *is* untrustworthy," since I have no objective reasons I could give in support of such an assertion.

Sometimes, the sole reason for basing an assertion that *p* on third-person evidence is one's recognition of the rules governing the norms of assertion, even when one is not at all moved by the evidence in question, not even in a third-personal way. When that is the case, there will, typically, be no Moorean paradox. Thus, suppose I have to write an encyclopedia entry on metaethics, and imagine further that I believe that my own view, say, moral realism, is correct. Plainly, I cannot assert in my entry that moral realism is true, no matter how firmly I believe this. I would have to say that there is a debate about the matter, and that the jury is still out. Thus, in one sense, my experience can be described using Moorean syntax, "The truth regarding the status of moral claims is unknown, but I don't believe this (because I believe I know the truth)." In another sense, however, there is no Moorean paradox at all: if I truly and firmly believe moral realism, and I am asserting that the jury is still out solely for the purpose of obeying the norms governing assertion (but I am secretly thinking that others just don't see the truth I see), then there is no paradox. These considerations suggest that in the cases we began with, the paradox does not automatically arise out of the conflict between the third-person evidence and the first-person seeming: the agents are, in addition, moved by *both* the public reasons and their own seemings.[29]

III.

There is a final point we wish to touch upon briefly before concluding this discussion. Many psychiatric patients—phobics, anorexics, delusional patients—assert what seem to be Moorean sentences. For instance, an arachnophobe may accept the evidence that some spiders are benign, yet fear those no less than toxic spiders. Thus, cleaning her basement one afternoon, Jill, the arachnophobe, may see a benign Huntsman spider and cringe in fear, thinking to herself, "Huntsman spiders are harmless, no need to fear them; but I don't believe that." Similarly, an anorexic person may accept the evidence showing that he is underweight, yet remain convinced that he is too fat. Chris the anorexic may remark, "Objectively, it is true that I am way too thin; but somehow I can't believe this. I still believe I am fat."

These cases appear to fit, at least to some considerable degree, the framework we used earlier to explain the Skywalk, Train, and Toilet cases. The psychiatric patients have sufficient third-person evidence—in the spider phobic case, that Huntsman spiders are not to be feared; and in the anorexic case, that a person, such as the patient himself, with a body mass index (BMI) of 17, is severely

[29] Of course, the encyclopedia case is not exactly parallel to the Train track and Skywalk cases, since the private seeming in the encyclopedia case is a result not of a-rational primary processes, but of reasoning. Nonetheless, the two cases are importantly analogous: in both, there is a conflict between what seems to me to be the case and what third-personal evidence supports.

underweight. Yet, at the same time, they may have a kind of seeming—so compelling that it gives rise to a paradoxical belief—in the untruth of the very proposition they endorse on the basis of third-person evidence. This precludes wholeheartedness in the endorsement of the proposition based on third-person evidence. So here is the puzzle. While in the Skywalk and Train cases discussed earlier, there is a-rationality in the way a seeming is generated, but no *ir*rationality in the belief formed on the basis of that seeming (this despite the fact that the first-person belief conflicts with the third-person evidence-grounded belief about what the world is like), there is indeed irrationality in the psychiatric patient cases. Why? Just what is the difference between the two kinds of cases?

Part of the answer has to do with the psychiatric patients' primary processes. Primary processes are operating in both sorts of cases, delivering experiences of "seeming fat", "seeming to be in danger", and the like, however, the processes are operating abnormally in the psychiatric patients. While a normal, rational person, given the potency of the associations, can be expected to be fearful on the Skywalk—and even more so under a train—however strong the third-person evidence of safety may be, a sane person is not expected to feel an incapacitating fear of spiders she has strong evidence to believe are benign, nor to form the clearly factually false belief that he is underweight. That it seems to the severely underweight patient that he is overweight indicates a failure of the primary processes to function normally—a patient with a BMI of 17, does not *look* overweight. So what gives rise to the seeming that he is? Our answer is that the seeming is rooted in a phantasy the patient fails to recognize as such. The patient may, for instance, be pre-occupied with a tiny amount of fat tissue he sees and feels on buckling his belt, but the patient's imagination blows this out of proportion. The anorexic's primary processes are, thus, inappropriately influenced by phantasy, leading to an abnormal "seeming." Otherwise put, one must be anorexic in order for it to seem to one that one's body with a BMI of 17 is fat.

This is not all, however. There is a second part: the psychiatric patients not only subjectively experience but miscategorize and treat their primary-process-driven beliefs as though they were secondary-process-mediated beliefs.[30] This leads to harmful, symptomatic, and irrational behaviors, but more importantly for present purposes, it signifies irrationality in the patient's belief system. If the patients didn't treat the primary-process-driven beliefs as secondary-process-mediated beliefs, they would not be irrational. For instance, if Chris, the anorexic, had a compelling seeming that he is fat which led to the Moorean "I am not fat, but I don't believe that," but he acted in accordance with the evidence rather than his own seeming, i.e., tried to gain weight rather than lose weight, Chris would no longer be irrational, although there would still be something not fully *sane* about him and his cognitive processes.

Similar considerations apply to the arachnophobe case. If our analysis is right, some arachnophobes are exhibiting irrationality while others are not. This is because some spider phobics recognize their intense phobic responses to spiders as automatic associations, based on a-rational processes operating abnormally, very different from the rational beliefs about benign spiders that they simultaneously hold. Other spider phobics believe all spiders are dangerous—linked inextricably in a-rational similarity assessments, aversive conditionings, and associations—rational knowledge of the benign Huntsman discounted. Only if Jill is in the first group will her Moorean assertions be neither absurd nor irrational.

---

[30] Though we cannot develop the point fully here, we note that the patients in question do not have stand-alone, objective, third-person, evidence-sensitive beliefs about the toxicity of Huntsman spiders nor the fatness associated with very low BMIs; in addition (and often instead) they have "neurotic beliefs." Briefly, neurotic beliefs are amalgams – composite propositions funded by evidence-insensitive, primary process-mediated unconscious phantasies, but treated as though they were beliefs-proper. See Brakel 2001 and Brakel 2009, Chapter 7. Neurotic beliefs are clearly quite similar to Gendler's aliefs in that they are both associational, a-rational, and automatic rather than rationally mediated, but they are different, as we state above, in that they are miscategorized as beliefs-proper and treated as such.

This brings us back to the point we made in the previous section—while a person may hold a Moorean belief based partly on a first-person seeming and partly on third-person evidence, and do so without irrationality, she can only maintain her rationality if she is properly sensitive to the nature of the grounds of her belief. A belief based on a first-person seeming ought not to be treated as a belief based on third-person evidence. (This is all the more true when the first-person seeming is based on a phantasy, rather than on primary processes that function normally.) In the previous section, however, we were concerned with the appropriate grounds of assertions: we suggested that a rational person is disposed to base her claims about the way the world is on objective evidence, not on her own seemings. We now wish to make an important addition: rationality requires that one be *sufficiently moved* by the third-person evidence, and not only in the sense that one's assertions about the world are based on that evidence, but in the sense that one acts on the evidence.

Are Luke and Kolya "sufficiently moved" by the third-person evidence, in this sense? Kolya clearly is: only a hyper-rational person can be so fully in control of himself and his actions as to be able to lie on the train track without fear. That Kolya is able to lie on the track at all signifies that he is moved by the third-person evidence more than an average person.

What about Luke? He is clearly *partly* moved by the third-person evidence: he allows his children to cross. Yet, he himself refuses to cross. Is this compatible with his being appropriately moved by the objective evidence? We wish to suggest that the answer is "yes." There isn't any strong reason for Luke to cross that would make it irrational for him to refuse. Not so with the anorexic patient: there is a very strong reason for the severely underweight patient to try to consume adequate calories: continued calorie restriction would be very harmful. Ignoring this strong evidence of harm is irrational.

This suggests that our earlier explanation of the rational assertibility of Moorean sentences we imagined Luke and Kolya Krasotkin uttering is incomplete. We first argued that the reason those sentences can be asserted without irrationality is that both parts of the Moorean conjunction are formed on the basis of evidence, and that the first-person seeming is so compelling that an ordinary person cannot be expected to simply form a belief against that evidence. We then said that this is not the full picture: rationality requires that a person be appropriately sensitive to the origin and nature of the evidence that grounds the two parts of the conjunction. We now see that these two claims together do not yet give us the full picture. Our discussion of psychiatric cases suggests that "appropriate sensitivity" here has a pragmatic element. While rationality does not require a person to be able to fearlessly lie on a train track or even cross a transparent bridge, it does require one to respond to strong evidence of harm. We believe that it is true more generally – and not only in cases involving Moore's paradox – that where one's beliefs furnish grounds for action, the rationality of one's beliefs is intimately connected to the rationality of one's behavior. This, however, is a topic for another paper.

IV.

It is typical to suppose that a Moorean sentence cannot be thought or uttered without absurdity and irrationality. If we are right, this is because the kinds of cases others have focused on do not recognize that the conflicting contradictory contents can arise, as in the cases discussed here, between substantially different cognitive processes—those that are associated with the a-rational System I, and those associated with the rational System II, systems that are functionally independent. Since the sorts of cases we discuss occur with considerable regularity, we believe that a failure to account for them constitutes a serious omission. Thus, we submit that, though in one sense this article is just another article on Moore's paradox, we don't believe that.

BIBLIOGRAPHY

Adler, Jonathan E., and Bradley Armour-Garb. 2007. "Moore's Paradox and Transparency," in Green & Williams 2007, 146-162.

Banks, Gordon; Short, Priscilla; Martinez, Julio; et al. 1989. "The Alien Hand Syndrome: Clinical and Postmortem Findings," *Archives of Neurology* 46: 456-9.

Borgoni, Cristina. 2015. "Dissonance and Moorean Propositions," *Dialectica* 69: 107-27.

Brakel, L.A.W. 2001. "Phantasies, Neurotic-beliefs, and Beliefs-Proper," *American Journal of Psychoanalysis* 61: 363-89.

Brakel, L.A.W. 2009. *Philosophy, Psychoanalysis, and the A-Rational Mind*. Oxford: Oxford University Press.

Brakel, L.A.W. & Shevrin, H. 2003. "Freud's Dual Process Theory and the Place of the A-Rational," *Behavioral and Brain Sciences* 23: 527-8.

Chislenko, Eugene. 2016. "Moore's Paradox and Akratic Belief," *Philosophy and Phenomenological Research* 92: 669-90.

Crimmins, Mark. 1992. "I Falsely Believe That P." *Analysis* 52: 191.

de Almeida, Claudio. 2007. "Moorean Absurdity: An Epistemological Analysis," in Green & Williams 2007, 53-75.

Dostoyevsky, Fyodor. 2002 [1880]. *The Brothers Karamazov*. New York, NY: Farrar, Strauss, and Giroux.

Fernández, Jordi. 2005. "Self-Knowledge, Rationality and Moore's Paradox," *Philosophy and Phenomenological Research* 71: 533-56.

Freud, Sigmund. 1953 [1900]. *The Interpretation of Dreams* (*Standard Edition* 4-5). London: Hogarth Press.

Gallois, André. 2007. "Consciousness, Reasons, and Moore's Paradox," in Green & Williams 2007, 165-88.

Gendler, Tamar. 2008. "Alief and Belief," *Journal of Philosophy* 105: 634–63.

Gertler, Brie. 2011. "Self-Knowledge and the Transparency of Belief," in A. Hatzimoysis, ed., *Self-Knowledge*. Oxford: Oxford University Press, 125-45.

Green, Mitchell and John N. Williams (eds.). 2007. *Moore's Paradox: New Essays on Belief, Rationality, and the First Person*. Oxford: Oxford University Press.

Heal, Jane. 1994. "Moore's Paradox: A Wittgensteinian Approach," *Mind* 103: 5-24.

Huemer, Michael. 2007. "Moore's Paradox and the Norm of Belief," in Susana Nuccetelli and Gary Seay, eds., *Themes from G.E. Moore: New Essays in Epistemology and Ethics*. New York: Oxford University Press, 142-57.

James, William. 1950 [1890]. *Principles of Psychology*, Vol. 1 & 2. Mineola, NY: Dover Publications.

Johnson-Laird, Philip. 1983. *Mental Models*. Cambridge, MA: Harvard University Press.

Kahneman, Daniel. 2011. *Thinking Fast and Slow*. New York, NY: Farrar, Strauss, and Giroux.

Kind, Amy. 2003. "Shoemaker, Self-Blindness and Moore's Paradox," *Philosophical Quarterly* 53: 39-48.

Kobes, Bernard. 1995. "Telic Higher-Order Thoughts and Moore's Paradox," *Philosophical Perspectives* 9: 291-312.

Kriegel, Uriah. 2004. "Moore's Paradox and the Structure of Conscious Belief," *Erkenntnis* 61: 99-121.

Larkin, William. 1999. "Shoemaker on Moore's Paradox and Self-Knowledge," *Philosophical Studies* 96: 239-52.

Lee, D. Byeong. 2001. "Moore's Paradox and Self-Ascribed Belief," *Erkenntnis* 55: 359-70.

Moore, G. E. 1944. "Russell's Theory of Descriptions," in P. Schilpp, ed., *The Philosophy of Bertrand Russell*. La Salle, IL: Open Court.

Moran, Richard. 2001. *Authority and Estrangement*. Princeton, NJ: Princeton University Press.

Niesser, Ulric. 1963. "The Multiplicity of Thought," *British Journal of Psychology* 54: 1–54.

Nussbaum, Martha. 2001. *Upheavals of Thought: The Intelligence of Emotions*. Cambridge: Cambridge University Press.

Piaget, Jean. 1926. *The Language and Thought of the Child*. London: Routledge & Kegan Paul.

Pruss, Alexander. 2012. "Sincerely Asserting What You Do Not Believe," *Australasian Journal of Philosophy* 90: 541-6.

Rosenthal, David. 2002. "Moore's Paradox and Crimmins's Case." *Analysis* 62: 167–71.

Rosenthal, David. 1986. "Two Concepts of Consciousness," *Philosophical Studies* 94: 329-59.

Schwitzgebel, Eric. 2010. "Acting Contrary to Our Professed Beliefs or the Gulf Between Occurrent Judgment and Dispositional Belief," *Pacific Philosophical Quarterly* 91: 531-53.

Shoemaker, Sydney. 1995. "Moore's Paradox and Self-Knowledge," *Philosophical Studies* 77: 211-28.

Solomon, Robert. 1980. "Emotions and Choice," in *Explaining Emotions*, edited by Amélie Rorty, Los Angeles: University of California Press, 251–81.

Sorensen, Roy. 2001. *Vagueness and Contradiction*. Oxford: Oxford University Press.

Stanovich, Keith & West, Richard. 2000. "Individual Differences in Reasoning: Implications for the Rationality Debate," *Behavioral and Brain Science* 23: 645–726.

Stoljar, Daniel & Hajék, Alan. 2001. "Crimmins, Gonzales, and Moore." *Analysis* 61: 208–13.

Turri, J. 2010. "Refutation by Elimination," *Analysis* 70: 35-9.

Vygotsky, Lev S. 1987 [1934]. "Thinking and Speech," in *The Collected Works of L.S. Vygotsky*, Vol. 1, edited by Robert W. Rieber & Aaron S. Carton. New York: Plenum Press.

Williams, John. 2006. "Moore's Paradoxes and Conscious Belief," *Philosophical Studies* 127: 383-414.

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

Wittgenstein, Ludwig. 1980. *Remarks on the Philosophy of Psychology*, Vol. 1. Oxford: Basil Blackwell.