Keith DeRose, Yale University
https://campuspress.yale.edu/keithderose/

What follows are

Chapter 2: Moorean Methodology: Was the Skeptic Doomed to Inevitable Defeat?

and

Appendix B: Experimental-Philosophy-Style Surveys on AI's First Premise

from my book, *The Appearance of Ignorance: Knowledge, Skepticism, and Context, Volume 2* (Oxford UP, 2017). I'm putting these together into one document, because in sections 2.7-2.99 (pp. 53-59) of Chapter 2, I evaluate the power/plausibility of the skeptical premise that I don't know that I'm not a brain-in-a-vat, making use of, among other things, results of some x-phi surveys Joshua Knobe and I ran on that premise, which results are reported in the brief (three-page) appendix.

Background: What precedes this material in the book is Chapter 1, a reprinting of my old paper, "Solving the Skeptical Problem" (SSP), *The Philosophical Review*, 1995.

"AI" refers to the Argument from Ignorance, the basic skeptical argument that goes:

1. I don't know that not-H.
2. If I don't know that not-H, then I don't know that O.
So, C. I don't know that O,

where "H" is a suitably chosen skeptical hypothesis (e.g., I am a bodiless brain in a vat who has been electrochemically stimulated to have precisely those sensory experiences I've had) and "O" is a proposition about the external world one would ordinarily think one knows (e.g., I have hands).

References:

Byrne, Alex 2004. 'How Hard are the Skeptical Paradoxes?', *Noûs* 38: 299-325.

Christensen, David 1993. 'Skeptical Problems, Semantical Solutions', *Philosophy and Phenomenological Research* 53: 301-321.

Conee, Earl 2001. 'Comments on Bill Lycan's Moore Against the New Skeptics', *Philosophical Studies* 103: 55–59.

DeRose, Keith 1999. 'Responding to Skepticism', in K. DeRose, & T. A. Warfield (eds.), *Skepticism: A Contemporary Reader*, New York: Oxford University Press, 1-24.

Foley, Richard 1993. *Working without a Net: A Study of Egocentric Epistemology*, New York: Oxford University Press.

Greco, John 2000. *Putting Skeptics in their Place*, Cambridge: Cambridge University Press.

Kelly, Thomas 2005. 'Moorean Facts and Belief Revision, Or Can the Skeptic Win?', *Philosophical Perspectives* 19: 179-209.

Kitcher, Philip 1992. ‘The Naturalists Return’, *Philosophical Review* 101: 53-114.

Klein, Peter 1985. ‘The Virtue of Inconsistency’, *Monist* 68: 105-135.

Kyburg, Henry. 1970. ‘Conjunctivitis’, in M. Swain (ed.), *Induction, Acceptance and Rational Belief*, New York: Humanities Press, 55-82.

Lewis, David 1996. ‘Elusive Knowledge’, *Australasian Journal of Philosophy* 74: 549-567.

Lycan, William G. 2001. 'Moore against the New Skeptics', *Philosophical Studies* 103: 35–53.

Moore, G. E. 1959b. ‘Four Forms of Skepticism’, in G.E. Moore, *Philosophical Papers*, London: George Allen & Unwin Ltd, 196-226.

_____ 1959c. ‘Certainty’, in G.E. Moore, *Philosophical Papers*, London: George Allen & Unwin Ltd, 227-251.

Nozick, Robert 1981. *Philosophical Explanations,* Cambridge, Mass.: Harvard University Press.

____ 1993. *The Nature of Rationality*, Princeton: Princeton University Press.

Pryor, James 2000. ‘The Skeptic and the Dogmatist’, *Noûs* 34: 517-549.

Roush, Sherrilyn 2010. ‘Closure on Skepticism’, *Journal of Philosophy* 107: 243-256.

Seyedsayamdost, Hamid 2014. ‘On Gender and Philosophical Intuition: Failure of Replication and Other Negative Results’, *Philosophical Psychology* 28: 642-673.

Sosa, Ernest 1999. 'How to Defeat Opposition to Moore', *Philosophical Perspectives* 13: 141–153.

Stroud, Barry 1984. *The Significance of Philosophical Scepticism*, Oxford: Oxford University Press.

Unger, Peter 1975. *Ignorance: A Case for Scepticism*, Oxford: Oxford University Press.

2

Moorean Methodology

Was the Skeptic Doomed to Inevitable Defeat?

2.1. Methodological vs. Substantive Mooreanism

My response to skepticism in Chapter 1 (“SSP”) is “Moorean” in a couple of different ways. First, it is *substantively* Moorean. Responses to AI can be classified into broad camps according to whether they deny AI’s first premise, deny its second premise, or accept its skeptical conclusion. Because G. E. Moore prominently walked this path, responses that deny AI’s first premise can be called “Moorean” responses. Relative to the BIV hypothesis, then, a substantively Moorean response involves the claim that one does indeed know that one is not a BIV. As we will see in a bit more detail in Chapter 3, though matters are muddied a bit by its contextualist character, my response, like that of most other contextualists, is in an important way “Moorean” in this substantive sense, because we hold that we know$_o$ that we are not BIVs: We know this by ordinary standards for knowledge. Ours can then be profitably classified as “contextualist Moorean” responses to AI.

A second way that SSP can be aptly called a “Moorean” treatment of skepticism is that it embodies a conservative, puzzling-solving approach to the problem of skepticism that is inspired by and was famously exemplified by Moore—though I expand on Moore’s approach in a way we will discuss. Though this approach animates all of SSP, it comes out most explicitly in its first and its last sections. In being “Moorean” in this *methodological* way, I again find myself at least roughly in league with the other contextualist Mooreans, but also with many other recent epistemologists who tackle skepticism.

My direct aim in SSP is to (a) “defeat” (b) the “bold skeptic” (c) who utilizes an argument from skeptical hypotheses like AI (though much of what I do has applications to skeptics who use different arguments). Chapter 4, on the contextualist nature of my solution, will be concerned with (b), explaining how the “bold skeptic” I target compares with other skeptics. The current chapter focuses on (a) and (c), primarily explaining, defending, and developing the broadly “Moorean” method of engagement with skepticism by which, as I will put it, one seeks to “defeat” the skeptic. I will also discuss the related questions of how important and powerful AI is, and whether the AI-wielding skeptic (and radical skeptics generally) had (has) any chance of "winning" a debate that follows conservative, Moorean, puzzle-solving methodological rules, or whether they were (are) doomed from the outset to inevitable defeat. In Chapter 3, we will take a comparative look at the alternative, "Refuting," method of engaging skepticism.

2.2. A Quick Look at Moore in Action

In "Four Forms of Scepticism" (Moore 1959b), Moore considers a skeptical argument of Bertrand Russell's, quite different from AI, to the conclusion that he does not know "that this is a pencil or that you are conscious."[1] After identifying and numbering four assumptions on which Russell's argument rests (the content of which we here ignore, so as to better focus on methodological matters), Moore writes:

> And what I can't help asking myself is this: Is it, in fact, as certain that all these four assumptions are true, as that I *do* know that this is a pencil and that you are conscious? I cannot help answering: It seems to me *more* certain that I *do* know that this is a pencil and that you are conscious, than that any single one of these four assumptions is true, let alone all four. That is to say, though, as I have said, I agree with Russell that (1), (2) and (3) *are* true; yet of no one even of these three do I feel *as* certain as that I do know for certain that this is a pencil. Nay more: I do not think it is *rational* to be as certain of any one of these four propositions, as of the proposition that I do know that this is a pencil. (Moore 1959b: 226)

One sentence later, Moore's essay comes to a close, and as that last sentence (which we will look at in Section 2.5) does not settle the matter, Moore doesn't really explain what conclusion the above observations are being put forward in the service of, but given what precedes the passage, and also what that closing sentence says, it is natural to suppose that Moore is explaining why he won't, and why he thinks he rationally should not, follow Russell's skeptical argument to its radical conclusion.

Moore took a similar approach to AI-like skeptical arguments, where his own substantive "Mooreanism" was displayed. Here he is in "Certainty" (Moore 1959c), responding to the dream argument:

[1] As the first quotation we are about to look at shows, Moore (quite unwisely, I believe) vacillates freely between knowing and knowing *for certain*, sometimes presenting the skeptical arguments as attempts to reach the conclusion that we don't know the things in question for certain, and sometimes as urging the conclusion that we don't know them. I will treat Moore as addressing the issue of knowledge, but the reader should be aware that Moore also took himself to be writing about certain knowledge—which he thought amounted to the same thing. Though this is far from obvious, I suspect that "knows" and "knows for certain" are used to express at least roughly the same range of relations between subjects and propositions—though it is plausible to suppose that the "bottom" of the range of "knows" cannot be reached by "knows for certain." (The alternative view would be that "for certain" adds something to the content that's distinct from what's ever expressed by "know" itself, and so takes us outside of the range of contents that can be exactly expressed by unadorned uses of "know(s).") But despite this similarity in expressive range, the two terms typically have (often importantly) different contents within most particular contexts, with "knows for certain" expressing a more demanding relation.

> I agree, therefore, with that part of this argument which asserts that if I don't know now that I'm not dreaming, it follows that I don't *know* that I am standing up, even if I both actually am and think that I am. But this first part of the argument is a consideration which cuts both ways. For, if it is true, it follows that it is also true that if I *do* know that I am standing up, then I do know that I am not dreaming. I can therefore just as well argue: since I do know that I'm standing up, it follows that I do know that I'm not dreaming; as my opponent can argue: since you don't know that you're not dreaming, it follows that you don't know that you're standing up. The one argument is just as good as the other, unless my opponent can give better reasons for asserting that I don't know that I'm not dreaming, than I can give for asserting that I do know that I am standing up. (Moore 1959c: 247)

In this famous example of "reversing the argument" ("one person's *modus ponens* is another's *modus tollens*," as it's often enough said), Moore agrees with AI's second premise (at least in its dream argument form), but rather than joining the skeptic in then reasoning from the first premise to the skeptic's conclusion, Moore proposes a counterargument in which he holds fast to the claim that he does know that he's standing up, and uses that, along with the agreed-upon second premise, to reach the conclusion that he does know that he's not dreaming. Thus, while the skeptic argues "1; 2; therefore, C," Moore counters, "not-C; 2; therefore, not-1." Moore here cautiously claims that, unless and until the skeptic can come up with some new good reasons for going her way, "the one argument is just as good as the other," but we have good reason to suppose Moore thought that his argument was actually better, and that here, as with his similar response to Russell's skeptical argument, it is more rational to follow Moore in concluding a premise in the skeptic's argument is false than it is to follow the skeptic's argument to the conclusion to which it leads.

2.3. Conservatism and Making a "Moorean Choice"

Moore's response to skepticism certainly smacks of some kind of conservatism: If one is inclined to reverse arguments in this way, then since any argument for a strongly enough counter-intuitive conclusion will be subject to such a maneuver, one will be generally quite resistant to changing one's views in response to arguments. But this seems an unproblematic form of conservatism on display here: One *should* be resistant to changing one's views in the way at issue, it seems.[2]

One worry it is easy to have about Moore's procedure is that Moore *begs the question* against the skeptic. Moore shows little interest here in what we might call "refuting" the skeptic, by which we will mean: deriving an anti-skeptical result (when dealing with skeptics who traffic in AI, that result is usually that one knows that O, or that O is true, or at least that H is false) from argumentative starting points that do not beg the question against the skeptic, but are rather things the skeptic does, or in

[2] I originally defended Moore along the lines about to follow in DeRose (1999: 4–6). For a kindred exposition and defense of Moorean methodology along these lines, see Lycan (2001: esp. 39–40).

some strong sense, must (since she is committed to them) accept. In Chapter 3, we will consider how important an anti-skeptical task it is to in that way refute skepticism. But we are now considering Moore's quite different way of countering the skeptic's argument that does not limit itself to such non-question-begging starting points. We will call such a Moorean response an attempt to "defeat" the skeptic. But what can such "defeat" amount to, and how can there be any value in an attempt to counter the skeptic that *does* engage in question-begging?

Whatever ends might or might not be promoted by a "refutation" of skepticism (as described above), it is Moore's non-refuting mode of response that answers directly to the important philosophical goal of responding rationally to arguments. From the standpoint of one seeking to govern her beliefs or acceptances rationally in response to reasons and arguments, one of the first and most important things to notice is that the question-begging between Moore and the skeptic is mutual, so "follow the argument that isn't question-begging" won't favor the skeptic. Yes, in utilizing the claim that he knows he's standing up as a premise of his counterargument, Moore certainly in some very good sense begs the question against the skeptic. The skeptic, after all, denies that premise, and has even offered an argument from intuitively plausible premises, AI, to back up that denial. But, by the same token, the skeptic's first premise is a claim that *Moore* denies, and Moore has offered an argument from intuitively plausible premises to back up *his* denial. Now, to those who are caught up in a certain kind of contest mentality, or to those who are interested in giving credit to the proper party for formulating a novel argument, it may be important whether the skeptic was here first, for, after all, in that case, Moore's argument is quite derivative. For certain purposes, one might even construe some debating vice one might call "begging the question" in such a way that in such a situation it sticks only to the Johnny-come-lately. But in the philosophically important ways, what question-begging there is here is mutual. One who is interested in rationally guiding her beliefs or acceptances in response to arguments will perceive that what AI and Moore's counterargument bring to light is a certain conflict of appearances: Each of 1, 2, and not-C can appear to be true, but, since they together form an inconsistent triad, it seems they can't all be true. While we may be in the skeptic's debt for bringing this conflict to our attention, that is no rational reason for resolving it in the way the skeptic favors.

Relative to the goal of rationally guiding her beliefs or acceptances in response to arguments, then, what *is* one to do if a powerful argument is presented toward a conclusion the negation of which one finds very plausible—plausible enough to (at least) rival the plausibility of the premises? That is: How should one respond to a conflict among claims all of which one finds plausible? In the quotations displayed in Section 2.2, Moore motions toward two suggestions.

In the last sentence of the passage I have quoted from "Certainty," Moore seems to suggest that one look for deeper positive reasons supporting the various pieces of the philosophical puzzle in question. Indeed, that sentence reads as if it's the prelude to

Moore's giving such positive reasons for his claim that he knows that O, and comparing these with the reasons the skeptic can come up with to support the first premise of AI. Of course, if that could be done, it could certainly help. But since each of the puzzle members will be plausible in its own right, it may be difficult to find arguments for them whose premises are even more certain than are the puzzle members themselves. And, in "Certainty," Moore doesn't deliver the goods he can seem to be promising. Immediately after the quotation, he asks, "What reasons can be given for saying that I don't know for certain that I'm not at this moment dreaming?" and he goes on to very critically examine what he considers the best positive support the skeptic could give for AI's first premise. But Moore somehow never seems to get around to offering any positive argument for his own premise that he knows that O in this essay, much less to subject such arguments to the type of scrutiny he leveled at the skeptic's argument. Instead, in practice, he just relies on the intuitive plausibility of his own claim—which raises the question: Why can't the skeptic likewise just rely on her premise's own intuitive plausibility? Why must the skeptic come up with a supporting positive argument? To get to my best guess as to how Moore would answer this, we must look to his other, more important, suggestion.

So, let us suppose that we can't, or at least so far haven't, come up with any helpful further positive support of the kind suggested above. What we face when we thus hit the argumentative rock bottom of the various possible positions is a set of claims, each of which is plausible, but which cannot all be true. If we want to have a consistent position, we'll want to reject, or at least suspend belief in, at least one of the members of that set.

(Methodological digression: We actually should not be too quick to assume that we should maintain consistent beliefs in such a situation. Especially where the members of the set all seem to have about the same, high degree of plausibility, and super-especially when there are quite a few of them (which does not apply to our current puzzle), the option of continuing to believe all of them, while, of course, realizing that they can't all be true, and so perhaps tempering the degree of one's belief in each, seems an attractive possibility, and several philosophers have given strong reasons in support of the rationality of sometimes holding sets of beliefs one knows to be inconsistent.[3] And of course, in such a puzzling conundrum, the strategy of simply remaining noncommittal on one or more of the claims clearly has its attractions, and indeed, relative to some goals, is just good sense. I recognize that holding inconsistent beliefs or remaining agnostic on one of the claims would often be the way to go relative to various rational goals which concern having the best picture of the world one can *right now*. The pressure to choose a consistent and complete (taking a position on each of the premises and the conclusions of the arguments in question)

[3] For some sustained defenses of the rationality of holding beliefs one knows to be inconsistent, see Kyburg (1970), Klein (1985), chapter 4 of Foley (1993), and Christensen (1993); and for a couple of quick defenses, see Kitcher (1992: 85) and Nozick (1993: 77–8).

package of claims to accept (even if not to really believe, which I suspect is usually beyond reach in most philosophically interesting cases) in these situations seems to me to come from some good philosophical methodology whose purpose is hopefully to promote a better picture down the road a bit: The best way for us to proceed, and hopefully to arrive eventually at the best picture available to us, may be for us to try out various consistent and complete packages of views, perhaps accept them, defend them, attack them, evaluate their comparative merits, etc.)

So, supposing we do want to reject one of the initially plausible but mutually inconsistent claims that constitute our puzzle (and accept its negation), we of course face the question: Which one? What Moore suggests in the first of the passages quoted in Section 2.2 is that one should reject the claim that is least certain to one. Relative to Russell's argument (which, recall, we're keeping in the abstract, without looking at the content of its four premises), Moore finds Russell's (4) to be the least certain, and so it's that member of the set he rejects—though it also seems fairly clear that (1), (2), and (3) are for him all ahead of the likes of "I know that this is a pencil" in the relatively-uncertain-and-so-subject-to-rejection line. And of course it's Moore's judgment that when it comes to AI, it's the skeptic's first premise that should be so rejected.

If no further progress on the problem can be made, then perhaps the best we can do by way of rationally responding to the skeptic's argument and the puzzle it presents us with is making such a "Moorean choice": We should reject that member of the set of mutually inconsistent but individually plausible claims that is the least plausible or seems least certain. Indeed, proceeding by making such a Moorean choice clearly *is* the way to go, given an assumption we are about to make explicit. Recall that we are supposing that we want to accept a consistent and complete position, where we either accept or reject (accepting its negation) each of the premises and the conclusion of the argument in question, and that we have reached each position's argumentative rock bottom: No further positive arguments from even more plausible deeper premises are currently available. Well, then, given that last feature of our situation, it would *seem* that the initial intuitive plausibility of the various claims is all they can have going for them, on the basis of which we might rationally decide which to accept. Given that assumption (which we will take back in Section 2.5), we should reject the least plausible of the mutually inconsistent claims. Better to reject what seems less plausible to us than what seems more plausible, if the claims' plausibility is all we have to go by. This would mean more generally that when the negation of a valid argument's conclusion is more certain or plausible to one than is one of its (needed) premises, one should "reverse the argument" and reject that premise before accepting the conclusion.

This understanding can also explain why, in our passage from "Certainty," Moore thought the burden was peculiarly on the skeptic to provide further support for her position: Since Moore judged one of the skeptic's premises (the skeptic's first premise, in the case of AI) to be the least certain of the conflicting claims, that was the claim that was first in line for rejection, unless it could be buttressed by a supporting argument. Unable to locate on the skeptic's behalf the support needed to get that claim pushed back in the line, the need for further support for the piece of the puzzle that Moore, but not the skeptic, was already inclined to hold on to (Moore's claim that he does know the relevant Os) never arose.

2.4. MORE PLAUSIBLE and its Application to the "Moorean Situation"

I am here accepting something like this principle that Thomas Kelly considers as one of several ways of fleshing out a (methodologically) Moorean position:

> MORE PLAUSIBLE: One should never abandon a belief in response to an argument when the proposition believed is more plausible than (at least one of) the premises of the argument.

But I accept something in the vicinity of MORE PLAUSIBLE only on a certain understanding of the "plausible" in contains. Kelly himself rejects the principle, at least when "plausible" is used in what he calls "its literal sense":

> Unfortunately for the Moorean, MORE PLAUSIBLE is false—at least, it's false if we understand "plausibility" in its literal sense. For strictly speaking, the plausibility of a proposition concerns, not its all-things-considered worthiness of belief, but rather its apparent or seeming worthiness of belief, or its worthiness of belief upon preliminary examination. Roughly: a proposition is plausible to the extent that it seems to be true to one who considers it. However, as Earl Conee has noted (2001: 57), plausibility in this sense is not a good candidate for being that which determines normative facts about what one ought to believe all things considered. Indeed, a given proposition's being extremely plausible is consistent with its being known to be false: Frege's Unrestricted Comprehension Principle does not cease to be plausible when one learns of its falsity. Given that plausibility is consistent with known falsity, it's clear that comparative plausibility is not the correct guide to belief revision. (Kelly 2005: 189)

But I disagree with Kelly about the literal meaning of "plausible":[4] I think we can use that word in a good and perfectly literal sense to designate plausibility in light of all the relevant considerations that we have with respect to a claim. (In this fine sense, Frege's Unrestricted Comprehension Principle becomes very implausible indeed when one learns that it is false: "'Implausible?' Yes! In fact, it's clearly false!")

[4] As will emerge in Section 2.6, I take myself to be largely in agreement with Kelly about the relevant Moorean methodology (at least until we get to my suggested improvement on that methodology, which Kelly hasn't considered), and our difference over MORE PLAUSIBLE seems to be a verbal one, generated by differences over the meaning of "plausible."

Where "plausible" is so understood—in what we might call its "all-in" use[5]—MORE PLAUSIBLE seems quite plausible, and it is not refuted by Kelly's argument.

Our application of the principle is, for now (but see note 7, where we update our application) to what we can think of as the "Moorean situation" where we have reached a case's "ultimate premises": There are no deeper arguments on offer in support of the premises in front of us. In such a case, all the premises *seem* to have going for them is something like their initial or intuitive plausibility—their apparent "worthiness of belief upon preliminary examination," as Kelly puts it. So, *in this situation* (though not generally), their all-things-considered plausibility would line up with their initial intuitive plausibility. Where all a needed premise really has going for it is its intuitive initial plausibility, it's hard to see how the rational response to the argument is to accept the argument's conclusion, if the opposite of that conclusion has more intuitive plausibility than does that needed premise.

2.5. Damage-Control Conservatism: Making an "Enlightened Moorean Choice" and the Project of Defeating the Skeptic

Following Moore in rejecting the least certain or plausible of the conflicting claims seems (far) more sensible than simply rejecting the claim that was initially fingered as the one to be rejected by the person who happened to first notice the conflict and who formulated an argument to the denial of one of the claims, using the other claims as the premises of her argument. Still, as I pointed out in Section 1.1 of SSP, making a "Moorean choice," as we have so far described it, isn't very satisfying. If indeed no further progress could be made, that would be a sad result. For rejecting something on the grounds that other propositions one finds plausible imply its falsity is not very

[5] I think there are analogous uses of "appears" and other verbs of appearance ("looks," "seems," etc.) that are very important to the practice of philosophy and other intellectual endeavors. I think the best general approach to take is that claims of the form "It seems that p" report an impulse or push toward believing p that occurs at some stage of cognitive processing (where there is context variability in what stage one reports by use of such a claim), with the important "all-in" use being a special case, where one reports an impulse toward believing p relative to one's currently final stage of processing, where it's judged with respect to all the relevant considerations one has at one's disposal. Thus, even after being fully convinced that the two lines one is talking about are the same length (relative to the standard of precision one is employing), one can still truthfully report that "The one on the left [looks, appears, seems] longer," when faced with what one knows is the Müller-Lyer illusion, because, though the impulse is completely shot down at some later stage of processing, one can still report the impulse toward believing that the line on the left is longer that is generated at some stage of cognitive processing. And when faced with some figures on a blackboard that one does not know to be a Müller-Lyer illusion, where the evidence is starting to point strongly but inconclusively toward the conclusion that what one is facing in indeed such an illusion, and so the lines are indeed the same length, one can truthfully say, using the "all-in" sense, "It seems that the lines are the same length"—though one could also instead use "seems" differently, and truthfully report that "The line on the left seems to be longer."

fulfilling when what one rejects is itself plausible—even if it's not quite as plausible as the claims one retains.

But what, then, is an inquirer to do? Well, there *might* be not much of value that can be done; we might be stuck with a sad result. But SSP embodies an approach that seeks an avenue of progress that might be available even if helpful positive support for the various claims that constitute a puzzle is not forthcoming. Even if no further positive arguments based on deeper reasons are to be had, we might nevertheless rationally have more to go by than just the various claims' initial plausibility. One can still hope for an *explanation* of how we fell into the puzzling conflict of intuitions in the first place—an explanation that may provide guidance on how to extricate ourselves from the trap. Perhaps we can explain how premises that together imply a conclusion we find so incredible can themselves seem so plausible to us. Such an explanation can take the form of explaining, for the member of the set that one seeks to deny, why it seems to us to be true, though it's in fact false. This, in a natural use of the phrase, would be a case of *explaining away* the plausibility of that claim,[6] and it could be a rationally helpful guide in our choice to reject the claim in question. The game then would not be one of producing more positive support for the aspects of one's position that are already plausible anyway (as we're supposing we can't do, anyway), so much as one of *damage control*: One seeks to provide a deflationary explanation for why we have the misleading intuition we have about the plausible statement that one chooses to deny—though in the hands of a contextualist, the intuitive costs may first be distributed among more than one of the claims before they are explained away or mitigated, as different readings of the claims are introduced.

Though contextualist accounts of the key terms in a puzzle may help in one's task of so solving a philosophical quandary, it is vital to note that the task of damage control does not have to utilize any form of contextualism—and in SSP, we discussed some attempts at damage control for straightforward (non-contextualist) solutions.

Indeed, we here are following the lead of some natural defenses of skepticism, whose devisers have seemed to sense skepticism's need for such damage control, and have therefore supplemented the skeptical arguments with explanations of what makes us mistakenly think we know the things the skeptic argues we don't know. Skepticism is often accompanied by suggestions or hints to the effect that while we do not really know the items in question, we do know them *for practical purposes*, or know them *for current intents and purposes*, or know them *given certain assumptions*, and/or that it's in some sense *appropriate* or *useful for us to claim to know them*, or something along those lines. It is then further suggested that it's because we confuse our standing in one of these other relations to the items in question for our knowing

[6] Here, in stressing the role of explanation, my way of engaging skepticism bears an important resemblance to Nozick's—and was no doubt influenced (for the good, I hope) by Nozick's treatment. See Sections 7.1 and 7.2 for a comparison of my use of explanation in dealing with skepticism with Nozick's.

them that we mistakenly think we know what we in fact do not. A crucial part of the case of SSP is arguing that such explanations proposed on behalf of the skeptic don't succeed (Sections 1.15 and 1.16), but the instinct behind these attempts is sound. Since the skeptic is asking us to reject a claim with a good deal of intuitive power, it will be difficult for her arguments to have enough intuitive oomph to do the job. Sensing this, skeptics and their defenders have sought deflationary explanations for why we have the (misleading, according to them) anti-skeptical intuitions we have. Such explanations, if successful, could have combined with the intuitive power of the skeptic's premises to provide a successful skeptical strategy.

Instead, I have argued, we can better explain how we came to be in this intellectual predicament in the way presented in SSP. If so, we can say that we have *defeated* the skeptic who wields AI in support of bold skepticism. That is, we will have successfully made the case that the best available resolution of this puzzle of conflicting intuitions is not that of the bold skeptic.

But the "success" here is for each reader to judge. In the end, one still has to make a choice: An "enlightened Moorean choice," we can call it. We have only delayed, and hopefully enlightened, the choice we have to make, by comparing our initially distasteful alternatives in terms of implausibility-given-damage-control, rather than in terms of initial implausibility of bare denial.[7] On the matter of how one can rationally make such a judgment about relative plausibility or certainty, I'm at as much of a loss as Moore was in terms of helpful general advice. Here is the closing sentence of "Four Forms of Scepticism," which immediately follows what I quoted at the start of Section 2.4:

> And how on earth is it to be decided which of the two things it is *rational* to be most certain of? (Moore 1959b: 226)

So you must employ your own best judgment in making your choice. And that's still the case in making my proposed "enlightened Moorean choice." And my claim to have "defeated" the AI skeptic turns on just such a judgment call. But there is good reason to suspect that, after attempts at damage control have been registered, most people will judge that they are, and are rationally, less certain of at least one of the skeptic's premises than they are that they know at least some things about the external world, and so, following this neo-Moorean procedure, will correctly judge it rational *not* to radically revise their opinion about the extent of their knowledge in response to the skeptic's argument. This neo-Moorean game is still governed by conservative rules.

[7] Updating our understanding of MORE PLAUSIBLE (see Section 2.3), we still understand the occurrences of "plausible" in it as designating "all in" plausibility, but we are now applying the principle to a different state of the inquiry, in which we are evaluating solutions in light of the attempts at intuitive damage control that have been applied to the claims constitutive of the puzzle, and so are no longer limited to the initial intuitive plausibility enjoyed by the claims.

Since skeptics typically propose quite radical revisions in our assessments about what we know (or are justified, or rational, etc., in believing), conservative rules work against them, and their friends may object. But it's unclear what grounds they have for so objecting—unless it's that these rules violate their sense of what makes for a fair or an exciting contest whose result was unpredictable. For, as I stressed in Section 1.17 of SSP, in the best case, what the skeptic has is an argument from deeply felt intuitions of ours to her skeptical conclusion (together perhaps with a plausible-looking deflationary account of why it might misleadingly seem to us as if we do know). How then can it possibly be illegitimate to point out that other of our deeply held beliefs militate against her conclusion? And why should we give credence just to those of our beliefs that favor the skeptic? And if we can show that those beliefs that seem to favor the skeptic's solution can be accommodated or explained away in a non-skeptical solution better than the skeptic can accommodate or explain away our beliefs that are hostile to her, then we will have shown that the skeptic's is not the best or most rational resolution to the puzzle for us to adopt. We will have *defeated* this skeptic—which is my claim for what is accomplished in SSP.

2.6. Was the Skeptic Doomed to Defeat?

But was this skeptic doomed to inevitable failure in a way that we should have seen from the beginning, and are skeptics—or at least sufficiently aggressive skeptics—generally doomed to inevitable defeat, whichever of the basic types of skeptical argument they utilize? That of course depends in large part on the power of their arguments—a matter over which there seems to be a lot of disagreement.

In *The Significance of Philosophical Scepticism*, Barry Stroud describes one common reaction to arguments by skeptical hypotheses as follows:

> I think that when we first encounter the sceptical reasoning outlined in the previous chapter we find it immediately gripping. It appeals to something deep in our nature and seems to raise a real problem about the human condition. (Stroud 1984: 39)

The "sceptical reasoning" to which Stroud refers is his own rendition of Descartes's dream argument, from the first chapter of *Significance*, which works a bit differently from our formulation of AI, but is like AI in being an argument by skeptical hypothesis. Still, Stroud's observation applies to AI as well. Similarly, writing about a simple form of argument much like our formulation of AI, Peter Unger writes:

> These arguments are exceedingly compelling. They tend to make sceptics of us all if only for a brief while. (Unger 1975: 9)

When arguments by skeptical hypotheses are first presented to students in philosophy classes, *some* do have reactions roughly like those that Stroud and Unger describe, I have found. But many have a very different reaction, claiming to find the arguments far-fetched, ridiculously weak, and quite unthreatening; such a reaction is often accompanied by an exclamation along the lines of, "Aw, come *on*!"[8] Those inclined to react in this latter way may have rolled their eyes when at the opening of SSP (Chapter 1) I described AI as "powerful," and then may have grown increasingly impatient at the respect with which I continued to treat the argument. These differences in initial reactions of students are mirrored by a similar division in the attitudes of philosophers toward skeptical arguments like AI, with some experiencing them as deeply threatening, and others seeming to find them too weak to get worked up over. (My sense is that there are now more of the latter kind of philosopher and fewer of the former kind than there were in the more brooding days in the immediate aftermath of Stroud's *Significance*.) The latter tend to view the skeptic as doomed to failure from the get-go.

To make sense of the dismissive attitude, at least insofar as it is directed toward AI, I think a key distinction is needed, for that argument really is quite powerful, and is certainly not absurdly weak. The argument's premises do imply its conclusion, and each of its premises, considered on its own, enjoys a great deal of intuitive support.[9] The reaction that AI is weak is probably best refined to the actually plausible claim that, though the argument may be fairly strong (in terms of the intuitive plausibility of its premises), at least so far as philosophical arguments go, it is not strong *enough* to adequately support such a counter-intuitive conclusion as the one it bears. And the reaction that the skeptical argument is *absurdly* weak is probably best refined to the (actually plausible) claim that it is *nowhere near* strong enough to support such a counter-intuitive conclusion, due to the advisability of a Moorean reversal of the argument. The dismissive may be sensing that (and dismissive philosophers might be more-or-less explicitly thinking that) our knowing such things as that we have hands is, and perhaps is *clearly*, as my fellow contextualist Moorean David Lewis nicely puts this important conservative insight, "a Moorean fact. . . . It is one of those things that we know better than we know the premises of any philosophical argument to the contrary" (Lewis 1996: 549). This would still make sense of the objector's sense that the argument constitutes no real threat to establish its conclusion.

Is skepticism then a real threat? Kelly nicely expresses his own in-advance confidence that the skeptic is doomed (where his helpful note specifying which skeptics are being said to be hopeless is put in brackets below at the point in his text at which he attached the note) in this passage from his appropriately titled "Moorean Facts and Belief Revision, or Can the Skeptic Win?":

[8] I have come to suspect that the reactions to these arguments that philosophers get from their students depends a lot on the manner in which they are presented, and that those who themselves take these arguments to pose a serious and important threat will tend to inspire responses that are very sympathetic to the arguments. By contrast, where students sense that it is in a way safe, or perhaps somehow intellectually respectable, to be dismissive of the arguments, such dismissive responses will be much more common.

[9] We will discuss the intuitive power of AI's first premise, which I take to be the argument's weak link, in Sections 2.7–2.10; see note 12 for my estimation of AI's second premise, and directions to related discussion.

> My own sympathies lie with the Moorean. I believe that there are very substantial limits on how radical a change in our views philosophy might legitimately inspire. For example, in epistemology—the domain on which I'll focus in what follows—I suspect that, ultimately, the skeptic simply cannot win. [Here and below, I use "skepticism" generically, to refer to any sufficiently radical variety of the view (as opposed to, say, skepticism about the existence of God or about the claims of psychical research). If more specificity is wanted, one might take the claims of the Moorean as being directed at skepticism about our knowledge of the external world.] The sense in which the skeptic cannot win is not that he will inevitably fail to persuade us of his conclusion—that, after all, might be a matter of mere psychological stubbornness on our part, which would, I think, be of rather limited philosophical interest. Rather, the sense in which the skeptic cannot win is that it would never be reasonable to be persuaded by the skeptic's argument. Moreover, I think that this is something that we can know even in advance of attending to the specifics of the skeptic's argument: in a sense, the skeptic has lost before the game begins. (Kelly 2005: 181)

Given his talk of "specifics," let's not read Kelly as making any predictions concerning any completely new skeptical arguments that might come along and blindside us (though he may also have something so general in mind), but rather as urging something like this: We all have a pretty good idea of the kinds of philosophical arguments for skepticism that have been tried, and though there may be many different ways of trying to run them, and though we maybe haven't yet hit upon the very best way of working out those details, Kelly is expressing confidence that (well, he's saying he *knows* that, but perhaps we should tone that down a bit to get a more interesting question) nothing like *that*—no argument that is a way of working out the details on the kinds of skeptical arguments with which we are familiar—is going to succeed in making it reasonable for us to accept on its basis that we have no knowledge of the external world. Any such valid argument to one of the "sufficiently radical" skeptical conclusions will contain at least one premise that it would be more reasonable for us to reject, in a Moore-like fashion, than it would be for us to accept the radical skeptical conclusion.

This is interesting in large part because, while it nicely expresses a basic attitude toward skepticism that Kelly shares with many other philosophers, still other philosophers feel very differently. Here are some factors that might account for the differences in "in advance" attitudes toward the skeptic's chances. First, some might take much more seriously than Kelly does the thought that knowledge is (and that the skeptical arguments are revealing it to be) an *extremely* demanding relation—so demanding that we have alarming little of it (and none of it with respect to the external world). Second, others might find it more believable than Kelly does that we are (and that the skeptical arguments reveal us to be) in a *deplorable* epistemic condition with respect to many of our ordinary beliefs (including perhaps all of our external world beliefs). For this second group, it's not that knowledge is so demanding, but that our epistemic position is, or at least might well be, so utterly pathetic. They simply do not find this thought absurd, or perhaps even particularly implausible, especially when they entertain it in the light of what they consider to be a powerful skeptical argument. These in the second group are the people most likely to find skeptical arguments to represent a truly *menacing* threat and who react to the arguments along the lines that Stroud describes in the passage I quote at the start of this section.

Third, those who find themselves in either (or both) of the groups described above might be explicitly or implicitly accepting, or taking seriously, or just more seriously than Kelly might, some damage-controlling explanation put forward by skeptics or on the behalf of skeptics. As I urged in Section 2.5, the skeptic's chances don't depend wholly on the initial intuitive power of her premises and the degree to which her conclusion initially seems absurd, but also on the effectiveness of her attempts to explain away our pro-knowledge intuitions that are hostile to her. Kelly does not explicitly consider the possibility of damage control: He seems to be considering just a "Moorean choice" in the style of Moore himself, not an "enlightened Moorean choice" of the type we considered in Section 2.5. So it's hard to say whether Kelly would continue to find the skeptic's case so hopeless if he were to upgrade his philosophical methodology and take into account all of the skeptic's potential resources. That changes our question to "Can any form of the familiar skeptical arguments, *together with any form of the familiar skeptical attempts at damage-controlling explanations*, succeed?" where we understand "succeed" in terms of making it rational for us to accept the argument's skeptical conclusion. Perhaps Kelly would still take the skeptic to be doomed, but others might find the skeptic's prospects for successful-enough damage control to be significant.

Those who harbor a suspicion that knowledge might turn out to be extremely demanding might find promising skeptical explanations according to which we're often *close enough* to knowing (for practical intents and purposes; or, often enough, for then-current intents and purposes; or something along those lines) to make it in some way alright, or at least unsurprising, that we would speak and think of ourselves as knowing when in fact we don't know. Those who take skepticism to be a truly menacing threat might be more inclined toward explanations according to which the relation we really stand in with respect to those things we mistakenly take ourselves to know is something like *knowing given certain assumptions*. (If we are in a *deplorable* epistemic position with respect to those assumptions, then *knowing given those assumptions* might be very far indeed from knowing.) Of course, there are different ways of working out such explanations, but for our current purposes it is best to leave them quite vague, for the question we are currently considering concerns the *in-advance* prospects for *any* attempt at damage control of these rough types working for the skeptic.

For the record, I should come clean about my own—admittedly boringly moderate—in-advance attitude toward the skeptic's chances. I've always found the thought that knowledge might turn out to be extremely demanding (much more demanding than a lot of our ordinary, rather breezy, knowledge-ascribing behavior would on the surface seem to indicate) to be itself quite believable, in advance. But that it should be *so* demanding that (and that our position should be such that) we shouldn't have any external world knowledge at all has always struck me as in-advance quite implausible: It would take an unusually powerful philosophical argument to establish such a daring skeptical conclusion. However, AI (and nearby variants of it) has always struck me as a powerful argument. I've never myself been inclined to actually judge it powerful enough to establish its conclusion (when it is aimed at exemplary bits of external world knowledge), but I have always thought it strong enough that, by means of it, together with some suitable damage control, the skeptic had enough of a chance of "winning" to make the question of whether she could win an interesting one. I didn't feel I could safely write off the prospects for the skeptic "winning" by such means until I saw, at least in broad outline, how the skeptical puzzle produced by AI could be solved along the lines explained in SSP, and the reasons (explained in Sections 1.15 and 1.16) why such a solution would be superior to the skeptic's solution.

2.7. A Division among Philosophers over the Intuitive Power of AI's First Premise

Some of the recent dissatisfaction with AI-like skeptical arguments has not just been based on Moorean methodological thoughts combined with a general modest view of the power of AI-like arguments, but has been focused on what is taken to be an *especially* weak link in such arguments: AI's first premise (the claim that one does not know that H is false—in the form we have been considering it, the skeptic's claim that one does not know that one is not a BIV). In fact, some years back, there was something of a little trend in some philosophical circles of dismissing AI and arguments like it as weak and unimportant, and perhaps even not worthy of much attention, based on that weakness,[10] and I suspect that this continues to be the prevailing attitude in at least some regions of epistemology. (Those who were underwhelmed by AI sometimes contrasted it with other philosophical arguments they thought were more powerful. In Appendix A, I address a couple of these comparisons.) As Kelly (whom we might as well use as an example, since we have already been discussing him) expresses this more focused complaint:

> I don't think that this is an especially strong argument compared to others which the skeptic might offer. In particular, I think that (1) [I don't know that I'm not a Brain-in-a-Vat (BIV)] is

[10] I have encountered this trend mostly in conversation, mostly from philosophers a bit younger than me, but for examples of it getting out in print, see Pryor (2000: 522), which I blame for the trend, and also Byrne (2004), both of whom I wrestle with in Appendix A, as well as Kelly, from whom we're about to hear, for some examples. From that same rough age cohort, though, Sherrilyn Roush takes a very different attitude, as we'll see in the next paragraph of the text. For an example of an epistemologist roughly my own age who isn't very impressed by AI's first premise, see Greco (2000: 52). Ernest Sosa, whom I will discuss a bit in Section 2.9, has always taken a healthfully critical attitude toward AI's first premise.

> an extremely strong claim to take as an unargued-for premise in an argument that is supposed to establish skepticism. If the skeptic simply asserts (1), then I think that the non-skeptic is well within her rights to simply decline to accept it. (Kelly 2005: 206)

Applying our distinction from (the fourth paragraph of) Section 2.6, it isn't clear whether Kelly is asserting that the skeptic's first premise is simply weak—perhaps too implausible for it to be a well-motivated project to even try to discern what accounts for its plausibility—or whether he is here (as we saw he is earlier in his essay) concerned with the question of whether the skeptic's argument really threatens to successfully establish what it is designed to show and is merely judging that, relative to that daunting project, the premise is not (nearly) strong *enough* to bear the great dialectical weight put on a premise in an argument designed to support such an extraordinary conclusion.[11]

But at any rate, I agree that these skeptical-about-skepticism philosophers have correctly identified AI's weakest link as its first premise.[12] But there seems plenty of room to think that this first premise is very strong for a weakest link in a philosophical argument for so startling a conclusion as AI's is, and I have always taken it that the traditional attitude toward arguments like AI—which attitude persisted throughout most of philosophy through the time of the little trend discussed above—is more in line with the estimates we saw from Stroud and Unger toward the start of Section 2.6: That these are powerful arguments that philosophers should reckon with. Nozick is the great example of a philosopher with respect for the skeptic here. Nozick seems very impressed, not only with the intuitive power of AI's first premise, but with the whole skeptical argument. He in fact thinks that AI's other, second premise is the argument's weak link, and, though he ends up denying (2), he thinks its plausibility is very strong indeed, likening the closure principle on which it can be based, in terms of "intuitive appeal," to a steamroller (Nozick 1981: 206). AI's first premise is stronger still in Nozick's eyes: That the skeptic is right here was something that he says we "deeply realize," to the point that attempts to show the skeptic is wrong on this point "leave us suspicious, strike us even as bad faith" (1981: 201).

[11] It seems that the reasons that "the non-skeptic is within her rights to simply decline to accept" this premise, according to Kelly, might include that it leads to such an implausible destination. But on the other hand, there is an indication in this quotation that there are other skeptical arguments that do better, and especially if the hint is that these others might really threaten to establish an extraordinary skeptical conclusion, one naturally wonders what those wondrous skeptical arguments might be. But it's unlikely that Kelly thinks that other skeptical arguments are that powerful, since he seems to think the skeptic (at least when suitably aggressive) is generally doomed to failure. That AI is rated sub-par even relative to other inevitable failures gives some reason to think Kelly intends some stronger and more dismissive criticism than just that this skeptic has no real hope of winning.

[12] AI's other (second) premise is underwritten by "closure" intuitions, and the important question concerning its skeptical power, at least to my thinking, is whether the closure principle for knowledge that we end up with once we have modified it to handle the problems that arise in trying to formulate it will still be strong enough to be of use to the skeptic. In Appendix D, I argue that it is.

Sherrilyn Roush joins Nozick in her attitude toward AI, writing: “to say that this is an intuitively compelling argument is an understatement” (2010: 243).[13]

For what it’s worth, my own initial reaction to AI and particularly to its first premise, as best I can recall, was in-between the two positions described above, but was perhaps a bit closer to the estimation by the critics of the argument than to that by Nozick: Though I could always feel a strong intuitive pull toward accepting this premise, I found it far from compelling, and I also felt a significant opposing intuitive push toward denying it. (However, it must be admitted that some of this “opposing push” may have just been the result of seeing where this premise led.) This, along with my quite similar attitude toward AI’s other premise, was expressed in Section 1.1 of SSP, where I wrote: “to be sure, the premises are only plausible, not compelling.” Still, I found the intuitive appeal of AI’s first premise to be not only strong, but more than strong enough to make it well worth accounting for that pull—to the point that I invested much effort in that project.

2.8. More Curiously Varying Responses to AI’s First Premise: Attempts to Ask Non-Philosophers

With such a divide in opinion among philosophers over the intuitive power of a claim, one naturally thinks to ask non-philosophers what they think, in the hope that their reactions, which are perhaps not swayed by current professional philosophical fashion, can give some guidance on how intuitively persuasive the claim really is. (And whether or not it’s natural, it is at least a quite *common* reaction in these days of experimental philosophy.) I’ve tried this a bit, in different ways, with extremely mixed results. Before trying to reach an evaluation of the power and usefulness of AI, I will here briefly convey what I’ve found.

I had long been asking students in a very informal way what they thought of AI’s first premise, and of the skeptical argument as a whole, and had come to the conclusion that students tended not to find that premise compelling, or even particularly plausible. However, I was finding some curious variation in what other philosophers reported was their own students’ opinions on the matter. Most curiously, different philosophy teachers of quite different persuasions were finding that their students agreed with them! So, I decided to ask in a slightly more disciplined way. In an attempt to ascertain students’ opinions in a way as independent as possible from my own influence and also from worries about where accepting the claim might lead, at the beginning of several semesters, I presented students in a largish introductory philosophy class with the question of whether they did or did not know that they were not BIVs at first by itself, without the rest of the skeptical argument, and as

[13] Interestingly, both Roush and Kelly were Nozick’s students at Harvard. And, along with Nozick, Pryor—the philosopher I blame for this whole backlash against AI (see note 11)—was on Kelly’s dissertation committee at Harvard. On this issue, Kelly sided with Pryor, while Roush joined Nozick.

the first thing done in the first meeting of the class, so they would not know my own inclinations. (Of course, it is possible—indeed, quite plausible—that even when the premise is presented on its own, those to whom it is presented will sense the skeptical threat the claim poses to their knowledge of ordinary facts, and this may still color their reaction to the premise.) I had "I know that I'm not a BIV" and "I don't know that I'm not a BIV" written on the board. I prepared for the question by briefly explaining what a BIV is.[14] I then asked which statement (put in their own voice) they thought was right, asking them to clearly decide how they wanted to answer (so they wouldn't be influenced by their classmates), and took a show of hands.

The results were surprisingly (to me, initially shockingly) strong—in favor of the skeptic! I suspected the results were really about as strong as you can hope for from undergraduate classes on just about any question—including whether torturing and killing babies for the fun of it is wrong. Well over 80 percent of students (and not just of those who voted) answered that they didn't know, and in each case, less than five (often in classes of about one hundred students, which is what this course drew for about three of those years, at least for its opening meeting during Yale's "shopping period") voted that they did know. (So there were a few abstainers.)[15]

This was all very unscientific, and, of course, it is quite dicey to draw conclusions about the degree of the plausibility of a claim from results concerning what proportion of respondents choose that claim over its negation. Among other reasons for caution here is this: It could be that while a very strong majority will choose the claim, few of them find it a very clear matter. It could be something of a close call for almost everyone, but just a fairly close call that the strong majority are inclined to make one way rather than the other.

But these results did suggest—however uncertainly—that AI's first premise has a lot more intuitive power than I had been giving it credit for. Indeed, that its intuitive power *may* be great. Perhaps enough to vindicate even Nozick's attitude, and/or enough that we might properly call the conundrum AI presents us with a "paradox."

But one good reason for caution here is that you can get very different results when you ask non-philosophers about this matter in a different way. When I very recently got a bit more rigorous still, at least in some ways, and conducted (with Joshua

[14] It is essential to a fair posing of the question to refrain from using terms of epistemic appraisal in one's description of the hypothesis. So, for instance, one should not characterize what a BIV is by saying anything like: "so, a BIV *can't tell* whether it's experiencing a real world." But it is also essential to giving the idea of the hypothesis to convey that a BIV has sensory experiences just like it would if it were experiencing a real world (though I could certainly understand someone thinking even that characterization begs important questions about what sensory experiences are). I would use a rather minimal formulation like this (found in an old PowerPoint presentation for the start of an opening class meeting): "BIV: a bodiless brain in a vat, which is hooked up to a giant super-duper computer that, taking into account the motor output of the BIV, sees to it that the BIV is electro-chemically stimulated with sensory input exactly as if it were embodied and experiencing a real world."

[15] For more details about these informal surveys and their results, as well as some interesting discussion in the comments about them, see the blog post at *Certain Doubts* where I reported them: "Polls Show that the Skeptic is Right," June 24, 2004, at: <http://certaindoubts.com/polls-show-that-the-skeptic-is-right/>.

Knobe's help) an experimental-philosophy-style survey on the matter, I got very different results indeed—results which would be to the liking of those philosophers who dismiss AI due to the weakness of premise 1. When I asked on a couple of surveys (726 participants in total) whether people thought they knew they were not BIVs, only 41 percent chose "I don't know that I'm not a BIV," while 59 percent chose "I know that I'm not a BIV." (For details about these surveys, see Appendix B.)

Among the factors that might account for these very different results are differences in the pool of respondents and in the circumstances under which they were asked the same question. My more informal surveys were of Yale undergraduates, signed up for an introductory-level philosophy class, while the recent survey was of ordinary adults over the age of eighteen.[16] Perhaps more important was the setting: Rather than anonymously and privately taking a survey, my students were publicly raising their hands and taking a position in front of their classmates, in a fairly large class in the somewhat charged environment of the opening meeting of a semester. And also in front of me. One potential explanation for the strong showing of "I don't know" in the classroom shows of hands may be the worry that some students may have had that I might ask them to explain and defend their answer, combined with some thought to the effect that it might be harder to defend a claim to know something (especially something so exotic) than an admission that one doesn't know it. And of course, expectations of what I would be like, based largely on my being a philosophy professor who was about to lead them through a tour of Descartes' *Meditations*, may have come into play, even though this was almost all of the students' first encounter with me. This need not have been a conscious decision students made to give an answer different from what they really thought in order to avoid trouble, but could have been some other kind of tendency to gravitate toward what seems a more easily defensible answer in settings where a defense might be publicly called for.

2.9. Assessment: The Intuitive Power of AI's First Premise

In assessing the intuitive power of AI's first premise, it is important to clearly distinguish it from some other things that are often said against the BIV hypothesis. One quite common response to AI is to ridicule the idea that one might be a BIV. And when philosophers have told me that their students reject AI's first premise, it has sometimes turned out, upon a little further questioning, that what their students really said that's leading them to this impression are such things as that the BIV hypothesis is ridiculous, stupid, crazy, far-fetched, not worth taking seriously, etc.

[16] Or whatever type of people tend to take such surveys. See note 1 of Appendix B for some demographic information.

Upon careful consideration, interpreting such exclamations about the BIV hypothesis as denials of AI's first premise in particular seems rash—and is often a matter of filtering fairly indistinct expressions of impatience with AI through one's own (substantively) Moorean inclinations. One could just as well (and perhaps this would be just a bit better) instead take such reactions to be denials of AI's *second* premise: as claims that, because the hypothesis is so far-fetched (or whatever), one *doesn't have to* know it is false in order to know such things as that one has hands. Best of all, to avoid over-reading, would be to just take these reactions to be rather unfocused expressions of dissatisfaction with AI. As such, the philosophical stance they may most nearly approximate is the *methodologically* Moorean position that no such argument, based on such an outlandish possibility, has any hope of overturning something so solid as that one knows that one has hands.

I am not at all dismissive of such dismissive reactions. In fact, as we will see in Chapter 7, the sense we have that the BIV hypothesis is ridiculous plays the key role in my account of how we come to know, at least by ordinary standards, that we're not BIVs. (So I, in an important way, incorporate this common reaction into my own substantively Moorean response to AI, recognizing that the reaction itself is not specific enough to be classified as one on which we do know that we're not BIVs.) But when evaluating the power of AI, in fairness to the skeptic, it is important to keep in mind that such exclamations about the BIV possibility are all very different from the claim that one *knows* the hypothesis to be false. And in my experience, many who will make such exclamations are more reluctant to make the bold claim that they know that the hypothesis is false. The "Argument from Non-Stupidity" (ANS) (the hypothesis that I'm a BIV is not stupid; if the hypothesis that I'm a BIV is not stupid, then I don't know that I have hands; so, I don't know that I have hands) is very different from, and, I'd say, considerably weaker than, AI, including at its first premise—as are similar skeptical Arguments from Non-Ridiculousness, from Non-Craziness, from Non-Far-Fetchedness, and from Worthiness of Consideration.

Though I have been curious about non-philosophers' opinions about AI, and its first premise in particular, and though the results of my efforts to ascertain this certainly could have affected how I approach the argument, my own interest in the argument was never seriously challenged by the results of my efforts. As I've reported, I was always quite conflicted about AI's first premise in my own thinking, feeling strong intuitive pushes toward both accepting and denying it. So long as it seemed that the first push might well be the one pushing toward the truth of the matter, AI was something of a threat. And my interest in the argument didn't even depend on it's being a real threat: Even if I had come to apply my methodological Mooreanism to the argument in the same way Kelly did, and concluded that the skeptic really had no chance, I still would have wanted to account for the appeal of AI's first premise—both in order to helpfully engage with those who felt it more strongly than me and felt more threatened by the argument than I did, but also in case that push was rooted in some general tendency that might also push us toward judgments—perhaps misleading ones—that we don't know other things, where it was not lined up against so strong a counter-push.

The main thing I learned from conversations with students (as opposed to just polling them) was that most felt both of these intuitive pushes—though different individuals felt them in varying strengths. And those who seemed to be subject to just one of the two forces in question divided at least *fairly* evenly between those inclined to think in a fairly unconflicted way that we do know that we are not BIVs and those who think in the same unconflicted way that we don't. That has always been more than enough to motivate me to try to account for the intuitive power of the skeptic's claim. My attempts to ascertain non-philosophers' assessments of what I take to be AI's weak link have if anything increased the extent to which I thought an account of that premise's appeal was desirable. That there are settings where that appeal seems very strong would tend to show that the premise (and then the argument of which it is the weakest link) *can be* very powerful, even if there are other settings where that appeal is quite diminished. (And perhaps when it is presented by an actual skeptic, or a teacher doing a good job of playing the role of a skeptic, its power is significantly greater still than it is when presented by a philosophy professor who is trying to be as neutral as possible about the matter.) In the end, how most people will weigh the intuitive pushes that they feel against each other and come down on the issue when asked in various ways to vote on it is not all that important[17]—though the evidence reveals that I've always been somewhat curious about that fairly unimportant matter!

2.10. Contextualist Mooreanism and the Intuitive Complexity Surrounding AI's First Premise

But then, of course, the tendency or push to judge that we *don't* know that we're not BIVs is not the only one we have to account for. As Ernest Sosa wisely points out:

> Consider, moreover, the need to explain how the skeptic's premise—that one does not know oneself not to be radically misled, etc.—is as plausible as it is. That requirement must be balanced by an equally relevant and stringent requirement: namely, that one explain how that premise is as implausible as it is. To many of us it just does not seem so uniformly plausible that one cannot be said correctly to know that one is not at this very moment being fed experiences while envatted. So the explanatory requirement is in fact rather more complex than might seem at first. And given the distribution of intuitions here, the contextualist and the Nozickean, et al., still owe us an explanation. (Sosa 1999: 147)

[17] I don't intend this as a general statement about the importance of non-philosophers' opinions about philosophical matters. It is instead a statement about the relative unimportance of differences in results in this particular situation, where it seems fairly clear that there are strong intuitive pushes in both directions on a matter, and what we are measuring is just how many people feel one push more strongly than another in a particular setting.

While we contextualists have been primarily focused on explaining the intuitive push toward agreeing with the AI skeptic's first premise, contextual Mooreanism seems well-positioned to account for the intuitive complexity that Sosa rightly points to here. And indeed, since I am myself among those to whom it has never been "uniformly plausible" that we don't know that we're not BIVs (plausible, I suppose; but certainly far from uniformly so), it is perhaps unsurprising that I would end up at a position that holds good promise for dealing with this intuitive complexity.

Recall that on my position (as is the case with other "contextualist Moorean" accounts), we do indeed know by ordinary standards that we are not BIVs: We do know_o this. But we don't know_h that we're not BIVs: We don't know it by certain extraordinarily high standards. Now that "ignorance" of ours isn't unique to ~BIV: According to the contextualist Moorean we know little to nothing by standards h. The special reason we can seem to ourselves not to know the likes of ~BIV is that the standards by which we don't know that fact are the very ones that an attempt to claim such "knowledge" would have some tendency to put into play. It's because attempts to claim such "knowledge" have this tendency pushing them toward being false, and admissions that one does not "know" this have a push toward being true that it can so easily seem to us that we don't know that we're not BIVs. This appearance is enhanced, at least on my own account, by the fact that the tendency on which it is based holds for insensitive beliefs generally, so that insensitivity can come to seem to be a general indication of a lack of knowledge.[18]

But such a position seems very well suited for being able to also account for the opposite inclination—the tendency to think that, damn it all, one *does* know that one is not a BIV. For it contains the claim that we do know ~BIV by ordinary standards. Indeed, on my own picture (this wouldn't characterize contextualist Mooreanism in general), we're not only well-enough positioned with respect to *I am not a BIV* to meet ordinary standards for knowledge, but also to meet even most of the extraordinarily high standards that are sometimes in play—though of course not well-enough positioned to meet h (the standards that talk of BIVs has some tendency to put into play). In SSP (Chapter 1) I'm quite explicit about our being in as strong an epistemic position with respect to ~BIV as we are in with respect to *I have hands*,

[18] Alex Byrne (p.c.) has suggested to me that the initial plausibility he finds AI's first premise to have may be due, at least in his own case, to his implicitly relying on some bad, fancy reasoning that would yield that premise as a result. And we might wonder whether Byrne's speculation might be true of others, as well: Perhaps AI's premise is plausible, at least to many, because they implicitly and vaguely rely on some batches of reasoning that are made explicit in some "fancier arguments"—arguments that Byrne goes on to consider in sections 3–4 of Byrne (2004), and that he and I are very likely to agree turn out to be no good. Showing the power of the insensitivity account of that premise's plausibility, as I attempt briefly in SSP (Chapter 1), and at greater length in Chapter 6, should do much to address this interesting worry. I should perhaps add that if I were to be convinced that the insensitivity account of the plausibility of AI's first premise were completely off-track, I would consider the implicit-reliance-on-(likely bad)-fancy-reasoning hypothesis to be a very serious possibility. Indeed, even given the success of the insensitivity account, it seems likely enough that implicit-reliance-on-fancy-reasoning plays *some* subsidiary role in the extent to which at least some people find AI's first premise plausible.

which latter would seem to be something we're especially well-positioned to know. Of course, this surprising comparative fact can be due either to our being in a surprisingly weak position with respect to the latter (as the skeptic would have it), or to our being in a surprisingly strong position with respect to the former. The surprise, on my account, is how well-positioned we are with respect to our not being BIVs.

That we know$_o$ that we're not BIVs, as contextualist Mooreanism in general has it, would seem to have great potential to explain the intuitive push that many report (often in tension with a push in the opposite direction) to think they really do know that they're not BIVs. This is especially so when conjoined with the realization that the tendency toward standards h coming into play is one that the contextualist Moorean can (and I do) hold is a tendency that can be conversationally resisted. No wonder it can seem (at least to many) that they really do know this, even as there is (at least often) also some push toward thinking that they do not. The potential here is greater still for my own account, since on it, that we're not BIVs is one of the things we know the best, so to say: It is one of the things with respect to which we are in the strongest position to know. This might help somewhat to explain the vehemence with which some are inclined to claim to know that they're not BIVs, and can do much to make sense of such thoughts as: "Well, if I don't know *that*, what *can* I possibly know?!"

Of course, crucial to such a contextualist account of the intuitive complexity of this conundrum will be claims that varying standards for knowledge are affecting people's judgments here—often without the people involved being aware that the content of "know(s)" is so moving around on them. And of course, such contextualist claims are highly controversial, at least as of now. Whether the available evidence points toward or away from contextualism is something fought out at length in volume 1 (DeRose 2009). But those inclined to think in particular that it is highly implausible that such changes in content could occur without speakers being clearly aware of that happening should recall in particular a key response to that claim: Namely, that many speakers find it quite plausible that just such a thing is going on, and they come to judge that this often happens when speakers (like their own past selves) are not clearly aware of it going on, and it doesn't initially seem much more plausible to accuse *these* speakers to be wrong about the matter of whether such changes in meaning occur—to be blind to the context-*in*sensitivity of "know(s)"—than it is to accuse those who think differently about the matter of being blind to the context-sensitivity of "know(s)" (DeRose 2009: esp. 157–60).

2.11. The Value of AI, Whether or Not the Skeptic Had a Chance

By raising considerations in favor of the power of AI's weakest link, and by advocating for the relevance of damage control to a proper Moorean analysis of the skeptical predicament, I have been promoting a view on which AI is an initially powerful and potentially threatening philosophical argument—one whose potential impact on us, if we respond rationally to it, isn't clear from the outset. However, I realize that whether one takes such a view turns on certain judgment calls one has to make, and that different readers will make in quite different ways.

So it is important to emphasize in closing that not only does the value of studying skeptical arguments not depend on their being so powerful as to truly constitute "paradoxes," but that these arguments can be well worth studying even if one's initial reactions to them places one toward the extreme Kelly side of the continuum at which one judges them to be in-advance doomed. For even if an argument were not nearly powerful *enough* to have any hope of establishing its conclusion, if it were nonetheless even a fairly strong argument, it would still be a likely source of important information about the argument's subject matter—knowledge, in the case of AI. If initially plausible premises yield a strong skeptical conclusion—especially one strong enough that its being wrong is a "Moorean fact"—that would indicate that some premise we might otherwise be tempted to accept must be wrong, which is quite likely to point toward a kind of error we are (to at least some extent) prone to in thinking about knowledge. Indeed, a premise need not even be plausible in the sense of its being initially more plausible than not for it to be significant and potentially very helpful news that it is in fact false. If you're very much up in the air about a claim, feeling significant pulls toward both accepting and denying it, it would seem to be very significant to learn that denying it is the way to go. And even if you yourself feel no significant pull at all toward accepting it, you might still find it valuable to be able to show to others that it should be denied on the grounds that it leads to a sufficiently unacceptable conclusion. This all may explain why some who take these arguments to be doomed from the start do nevertheless devote much energy to them.[19]

Moore himself seemed to think he was learning something significant. In the case of each of the skeptical arguments he was addressing in the passages I quoted in Section 2.3, while Moore took the negation of the skeptic's conclusion to be considerably more plausible than the skeptical premise Moore chose to reject, he seemed in

[19] John Greco is a great example of this exemplary attitude. Though a central claim of Greco (2000) is that "a number of historically prominent skeptical arguments make no obvious mistake and therefore cannot easily be dismissed" (2000: 1), Greco does seem to take skepticism to be doomed to inevitable defeat in our sense, yet also thinks that it is very worth studying for what it reveals about its subject matter: "I argue that the analysis of skeptical arguments is philosophically useful and important. This is not because skepticism might be true and we need to assure ourselves that we know what we think we know. Neither is it because we need to persuade some other poor soul out of her skepticism. Rather skeptical arguments are useful and important because they drive progress in philosophy. They do this by highlighting plausible but mistaken assumptions about knowledge and evidence, and by showing us that those assumptions have consequences that are unacceptable. As a result we are forced to develop substantive and controversial positions in their place. On this view skeptical arguments are important not because they show that we do not have knowledge, but because they drive us to a better understanding of the knowledge we do have" (Greco 2000: 2–3). Unfortunately, because Greco is among those who fail to see the power of AI's first premise (see n. 11), AI is not one of the skeptical arguments that Greco applies his fine methodology to.

each case to think the premise he rejected was quite plausible, considered on its own. That it is false seemed to be significant news for Moore.

And, because of the explanations produced, when we advance to playing the "enlightened Moorean" game, we can increase what we can learn from strong skeptical arguments (even in the case where we can tell in advance that they have no hope of being strong *enough*), for we are now leveraging the pressure of avoiding implausibly strong skeptical claims to also rationally motivate the acceptance of claims about *how*, and not just *that*, interestingly plausible skeptical premises go wrong. And, at any rate, the intuitive power of a premise in a skeptical argument—like AI's first premise—needn't be all that great to make it enlightening to account for what intuitive pull it does have.

Furthermore, as we will see in Chapter 6, our tendency to accept AI's first premise is an instance of a very general tendency to judge that insensitive beliefs do not constitute knowledge. What we learn here will not be limited to some isolated intellectual glitch, but will point the way to an important lesson about knowledge and knowledge claims that is of very general application.

And, of course, the apparent importance of a skeptical argument only increases if one is instead among those who find it in advance to have a chance to work. In that case, not only the plausible-but-not-compelling premises of the argument, but also the not-so-Moorean-after-all negation of its conclusion, are all up for grabs. Finding the best solution to the puzzle that the argument presents us with will still promise to be a good way to investigate the concepts involved, but now one needs to determine not only how, but, first, whether, to evade the skeptical conclusion.

Appendix B: Experimental-Philosophy-Style Surveys on AI's First Premise

As reported in Section 2.8, when, with Joshua Knobe's help, I asked on a couple of experimental-philosophy-style surveys (726 participants in total) whether people thought they knew they were not BIVs, only 41 percent chose "I don't know that I'm not a BIV," while 59 percent chose "I know that I'm not a BIV." These numbers represent the results of two surveys, taken over three days.[1]

In the first survey, 215 participants were recruited using Amazon's Mechanical Turk and given this description:

> Let's use "BIV" (for brain in a vat) to mean a brain that has no body, but is being kept alive in a vat, and is hooked up to a super-advanced computer that sees to it that all aspects of a normal brain's interactions with its body and with the world around it are perfectly simulated. So everything seems to the BIV exactly as it would if it had a body and were experiencing an external world.

They were then asked:

> We would first like to ask the following question, which is designed to see if you understood what a BIV is.
>
> If a BIV, as just described, were having the experience of eating a blueberry, how would that seem to the BIV?

And almost all (94 percent) of respondents chose "It would seem to the BIV exactly as if it had a body and was eating a blueberry," with only 6 percent instead choosing the other option provided, "It would seem to the BIV just a little bit different from how eating a blueberry would seem to a normally embodied brain." On the main question, respondents were told "We are interested in whether you can know that you are not a BIV of the type just described," and then asked which of the two options they thought correctly described them, and 58 percent chose "I know that I'm not a BIV," and 42 percent chose "I don't know that I'm not a BIV." Thus, people's tendency to say that they knew they were not a BIV was significantly greater than what would be expected by chance alone, $\chi 2$ (1, N = 215) = 5.1, p = 0.02.

[1] I should note that my surveys drew quite a few more male than female respondents, which makes me wonder about that and also about potential other ways that the pool might not be representative of the general population. Of the 720 respondents who gave their gender (six didn't), 63 percent were male, while only 37 percent were female. I should note that there seemed to be a significant difference between the genders, with females being more inclined than males to respond that they do know that they're not BIVs: 68 percent of women said that they knew, while 54 percent of men said that they knew. This difference was statistically significant, $\chi 2$ (1, N = 720) = 13.4, p < 0.001. However, caution is advisable before concluding that there is a significant gender difference on this and other philosophical questions (see Seyedsayamdost 2014).

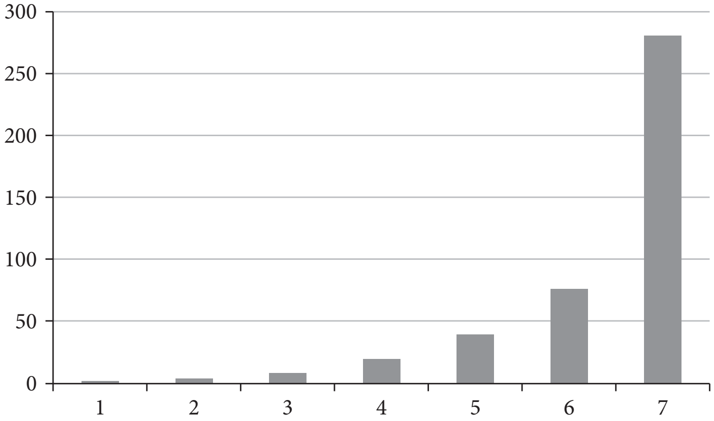

Figure B.1. Number of participants at each confidence level, among those who said that they did know they were not BIVs, collapsing across the two studies.

In a second survey, 511 participants were given this slightly different description:

> Let's use "BIV" (for brain in a vat) to mean a brain that has no body, but is being kept alive in a vat, and is hooked up to a super-advanced computer, that, taking into account the motor output of the BIV, gives the BIV appropriate sensory input. Because all aspects of a normal brain's interactions with its body and with the world around it are perfectly simulated in a BIV, everything seems to a BIV exactly as it would if it had a body and were experiencing an external world.

They were asked the same initial question as in the first survey, and this time 91 percent answered that things would seem to the BIV exactly as they would if it had a body and were eating a blueberry. On the main question, results were quite similar to the first survey: This time 60 percent chose "I know that I'm not a BIV" and 40 percent chose "I don't know that I'm not a BIV." Again, this is significantly greater than chance, χ2 (1, N = 511) = 20.8, p < 0.001.

As I remarked in Chapter 2, these are very different from the much more skeptic-friendly results I had earlier obtained by the quite different means of taking a show of hands among students in an introductory philosophy class. (See Section 2.8 for discussion of these differences.) This difference in results is rendered even more remarkable by the confidence that their answer was right that was reported by those who answered "I know that I'm not a BIV" on the later x-phi-style surveys. After answering the main question, I asked respondents, "How confident are you that your answer to the previous question is correct?" on a scale of 1 to 7, with 7 labeled as "most confident" and 1 as "least confident." Figure B.1 shows the distribution of confidence levels for participants who answered that they did know that they were not BIVs.

I find it arresting that on a question where there is a somewhat close 59 percent–41 percent split in answers, so many of those who gave the majority answer would be so confident that they are right.[2] (Recall, though, that each person is answering whether they take themselves to know that they're not a BIV. This opens the possibility—however slight—that most everybody

[2] The participants who said that they did not know (295 in total) that they were not BIVs were on the whole less confident that their answer was correct. The distribution was as follows: 36 percent (106) chose 7; 21 percent (63) chose 6; 18 percent (53) chose 5; 16 percent (48) chose 4; 4 percent (11) chose 3; 2

is right: Maybe many really do know this of themselves, while many others don't!) And this confidence of so many that they are right to think they know makes it seem, at least to me, even more remarkable that in the different setting of my asking students at the start of a philosophy course what they think, so very many would say that they do *not* know.

I have long been bothered by the confidence philosophers often project on our answers to questions we're in no position to know the answer to. (See Appendix C for related discussion.) These results tempt me to the thought that it's people generally who tend to be overconfident when they deal with philosophical questions—and that this afflicts philosophers more than others just because we spend more of our time on these matters.

percent (6) chose 2; and 1 percent (4) chose 1. Thus, there is a significant effect such that those who say that they do know show higher confidence levels, $t(719) = 8.0$, $p < 0.001$.