COULD KK BE OK?[*]

The KK principle—the principle that knowing entails knowing that one knows—has seen better days. In the past it was defended by Plato, Aristotle, Augustine, Averroës, Aquinas, Spinoza, Schopenhauer, and Prichard, to name a few.[1] Today it languishes in disrepute. Louise Antony refers to it as "the principle, roundly rejected by epistemologists of almost every stripe, that in order to know that *P*, I must know that I know that *P*."[2] Its fall is often thought of as closely linked to the rise of externalist theories of knowledge—theories that de-emphasize the role of the knower's reasons or justification for her beliefs (at least when reasons and justification are understood along traditional lines), and stress the importance of causal and/or nomological connections between the knower and the fact known.

Despite its current unpopularity among epistemologists, the KK principle—or at least the failure of popular objections to the KK principle—is often presupposed in work *outside* of epistemology. Much work in philosophy of language, game theory, and computer science appeals to the idea of "common knowledge."[3] The proposition

[1] Jaakko Hintikka, a staunch defender of the KK principle, mentions these and other prominent historical proponents. See his *Knowledge and Belief: An Introduction to the Logic of the Two Notions* (Ithaca: Cornell, 1962).

[2] Louise Antony, "A Naturalized Approach to the *A Priori*," *Philosophical Issues*, XIV, 1 (October 2004): 1–17.

[3] David Lewis's work on convention is perhaps the most influential example in philosophy. See David Lewis, *Convention: A Philosophical Study* (Cambridge: Harvard, 1969). See Peter Vanderschraaf and Giacomo Sillari, "Common Knowledge," *Stanford Encyclopedia of Philosophy* (Spring 2014 Edition), Edward N. Zalta, ed., URL = <http://plato.stanford.edu/archives/spr2014/entries/common-knowledge/>, for a helpful survey article on common knowledge that discusses, *inter alia*, uses of the concept of common knowledge outside of philosophy. It also provides some examples of how KK is often assumed in discussions of common knowledge.

© 2014 The Journal of Philosophy, Inc.

that *P* is common knowledge if not only does everybody know that *P*, but everybody knows that everybody knows that *P*, and everybody knows that everybody knows that everybody knows that *P*, and so on *ad infinitum*. But standard arguments against KK typically have the consequence that common knowledge is unattainable.[4] If these arguments are sound, we must reject explanations of linguistic, economic, and other social phenomena that appeal to common knowledge.[5] The debate over KK, then, has consequences that extend well beyond epistemology.

In this paper, I aim to push back against the turning tide of conventional wisdom.[6] Among epistemologists, both defenders and opponents of KK often assume that higher-order knowledge is a very exalted epistemic state. Knowing that one knows, or worse, knowing that one knows that one knows…*ad infinitum*, is agreed by both camps to be difficult or impossible for us mere mortals to achieve. Opponents of KK hold that we can have first-order knowledge without meeting the onerous requirements for higher-order knowledge, while defenders of KK hold that even first-order knowledge requires meeting those strict requirements; this is why the KK principle is often associated with skepticism.[7] The positive picture I will present—one fully compatible with the externalism commonly thought to be in conflict with KK—rejects the assumption often accepted by both camps; the picture that I will develop is one in which acquiring knowledge of arbitrarily high orders is no more difficult than we pre-theoretically take the acquisition of first-order knowledge to be. I motivate this picture by arguing that certain now-standard objections to KK deliver unhappy results when applied to examples in which higher-order *inter*personal knowledge—knowing

---

[4] I will explain in section I why these arguments rule out the possibility of common knowledge. Timothy Williamson, *Knowledge and Its Limits* (New York: Oxford, 2000), who argues against KK, embraces this result. He argues that the unattainability of common knowledge can be used to explain some otherwise puzzling phenomena. In my opinion these phenomena are best explained using other strategies.

[5] For example, Hawthorne and Magidor argue that failures of KK scuttle Stalnaker's account of assertion. See John Hawthorne and Ofra Magidor, "Assertion, Context, and Epistemic Accessibility," *Mind*, CXVIII, 470 (April 2009): 377–97.

[6] I defend KK (and BB) against some other objections in a companion paper to this one, Daniel Greco, "Iteration and Fragmentation," forthcoming in *Philosophy and Phenomenological Research*. For another recent defense of a version of KK, see Connor McHugh, "Self-Knowledge and the KK Principle," *Synthese*, CLXXIII, 3 (April 2010): 231–57.

[7] Barry Stroud, *The Significance of Philosophical Scepticism* (New York: Oxford, 1984), endorses KK, and appeals to it in offering certain skeptical arguments. Jonathan E. Adler, "Skepticism and Universalizability," this JOURNAL, LXXVIII, 3 (March 1981): 143–56, also holds that the KK thesis leads to skepticism, but rejects it on these grounds.

what *other* people know—is the focus. I go on to argue that the picture I present provides a more attractive account of such cases.

In section I I briefly describe two counterarguments to the KK principle. I then provide an example in which it is intuitively plausible that agents can gain higher-order interpersonal knowledge in certain ways. I argue that we cannot simultaneously accept (1) intuitive claims about what it would take for the agents in the example to gain higher-order interpersonal knowledge, (2) a closure principle concerning knowledge, and (3) the two counterarguments to the KK principle discussed at the beginning of section I.

I hope the example will have enough intuitive force on its own to suggest that something is wrong with standard arguments against KK. However, rather than simply resting on the intuitive force of the example, in section II I present a simple theory of knowledge on which closure and KK hold. I argue that it provides an attractive analysis of the example from section I, and that this reinforces the dialectical significance of that example. In section III, I argue that the same apparent counterexamples to KK that motivate its rejection are paralleled by apparent counterexamples to epistemic closure principles, and the same contextualist maneuvers that have been invoked to defend closure from counterexamples apply equally well in the case of KK.

I. TWO ARGUMENTS AND AN EXAMPLE

I.1. The Independent Check Argument. What does it take to know by vision that some object *x* is red? Plausibly, (at least if we are externalists of some sort) I must have a visual system that is capable of distinguishing red objects from non-red objects, and that system must function appropriately in some particular case to deliver me the information that some particular object is red.

Assuming I know by vision that *x* is red, what would it take for me to know that I know that *x* is red? According to what I will call the "independent check argument," I would need to avail myself of some *other* belief-forming process that could distinguish scenarios in which I have a reliable color-vision system from scenarios in which I do not. This could involve my having access to the testimony of an ophthalmologist who would have told me if my color vision were bad, or my having memories of situations in which I successfully exercised my color vision (memories that would have been different if my color vision were unreliable). Generally, knowing that *P* requires using some process that will reliably cause one to believe *P* only when *P* is true. Knowing that one knows that *P* requires having some *further* process that can reliably distinguish between

situations in which the process that produced one's belief that P is reliable, and situations in which it is not. Having higher-order knowledge requires that I be able to perform an independent check on whether my lower-order belief was reliably produced.[8]

Exactly how this argument works will depend on how we understand what it takes for a belief-forming process to be reliable, and what it takes for such a process to be able to distinguish between cases in which other processes are reliable and ones in which they are not.[9] However we flesh out the details, there will be cases in which something like the independent check argument seems to capture why we are happy to attribute knowledge to a subject, but not to attribute higher-order knowledge to that subject. For example, we might hold that a reliable chicken-sexer who has no information about her reliability can know whether a chicken is male or female, but cannot know *that* she knows whether the chicken is male or female, since she has no way of independently corroborating that her chicken-sexing faculty is reliable.[10]

I.2. The Safety Argument. A second argument against the KK principle appeals to the claim that if you know that P, then there is no nearby possibility in which you are wrong about whether P.[11] The claim is easy to motivate—it is plausible that if you could have easily been wrong about whether P, then you do not know that P.[12] If we can gloss "You could have easily been wrong about whether P" as "there is a nearby possibility in which you are wrong about whether P," then we get our claim that you cannot

---

[8] Many externalist/reliabilist denials of KK seem to be motivated by something like the independent check argument. It is natural to attribute a version of this argument to William P. Alston, "Level-Confusions in Epistemology," *Midwest Studies in Philosophy*, v, 1 (September 1980): 135–50; and Antony, *op. cit.*, also seems to have in mind an argument along these lines.

[9] This sort of argument is very naturally captured in a "sensitivity"-based framework of the sort defended by Robert Nozick, *Philosophical Explanations* (Cambridge: Harvard, 1981). For criticisms of that framework, see Williamson, *op. cit.*, chapter 7; and Jonathan Vogel, "Subjunctivitis," *Philosophical Studies*, cxxxiv, 1 (May 2007): 73–88. I do not assume, however, that the independent check argument can *only* be motivated if we assume a sensitivity constraint on knowledge.

[10] I do not know where chicken-sexing was first mentioned in an epistemological context, but Douglas Gasking, "Avowals," in R. J. Butler, ed., *Analytical Philosophy* (Oxford: Basil Blackwell, 1962), offers one early discussion.

[11] See Williamson, *op. cit.*

[12] This is a bit too quick; Nozick, *op. cit.*, p. 193, considers a case in which Jesse James is robbing a bank and his mask slips. A bystander sees James's face and thereby comes to know that Jesse James is robbing the bank, even though she could have very easily falsely believed otherwise (had the mask not slipped). Cases like these lead Nozick to opt for a more complicated version of the safety requirement, one that concerns the methods by which one forms beliefs. See footnote 50 for some brief discussion, and Williamson, *op. cit.*, chapter 7, for some extended discussion.

know that *P* if there is a nearby possibility in which you are wrong about whether *P*.

On this picture, a necessary condition for knowing that one knows that *P* is that there is no nearby possibility in which there is a nearby possibility in which one is wrong about whether *P*. The argument that this leads to failures of KK turns on the idea that the relevant nearness relation is not transitive.

A spatial analogy will help make this clear. There might be no nearby houses in which serial killers live, but there might be nearby houses in which there are nearby houses in which serial killers live (if, for example, houses count as nearby to one another iff they are within five blocks, and my house is seven blocks away from the closest house in which a serial killer lives). Similarly, there might be no nearby possibilities in which I am wrong about whether *P*, but there might be nearby possibilities in which there are nearby possibilities in which I am wrong about whether *P*. If things work out this way, then I may know that *P* while failing to know that I know that *P*. We might say that both cases involve being safe but not safely safe. I will call this argument against KK the "safety argument."

There are many other arguments against the KK principle. For instance, it seems possible to know that *P* without believing that one knows. If knowledge requires belief, then this entails that it is possible to know that *P* without knowing that one knows.[13] One way this might happen would involve creatures that do not even have the concept of knowledge—couldn't such creatures know things without even being *capable* of knowing that they know things?[14] Those sympathetic to the KK principle might try to meet these points head on.[15] But another strategy involves conceding the points and holding that there are still interesting versions of the KK principle that can survive this concession. For instance, one might retreat to the position that if someone knows that *P and* believes that she knows that *P*,

---

[13] The claim that knowledge requires belief is not uncontroversial. See Blake Myers-Schulz and Eric Schwitzgebel, "Knowing that P without Believing that P," *Noûs*, XLVII, 2 (June 2013): 371–84, for some recent criticisms. I am sympathetic to the defenses against these criticisms offered by David Rose and Jonathan Schaffer, "Knowledge Entails Dispositional Belief," *Philosophical Studies*, CLXVI, 1 Supplement (December 2013): 19–50; and Wesley Buckwalter, David Rose, and John Turri, "Belief through Thick and Thin," unpublished manuscript.

[14] Richard Feldman, "Fallibilism and Knowing that One Knows," *Philosophical Review*, XC, 2 (April 1981): 266–82.

[15] On the possibility of having higher-order knowledge while lacking the concept of knowledge, see the essays from the final section of Robert Stalnaker, *Context and Content* (New York: Oxford, 1999).

then she knows that she knows that *P*, or that if someone knows that *P*, then she is in some sense *in a position* to know that she knows that *P*.[16]

While I am sympathetic to a hard-line response, I will not attempt to rebut objections to KK that turn on the possibility of knowing without believing that one knows. For this reason, my defense of KK will be limited in scope.[17] But it will not thereby be deprived of interest. Part of what is interesting about the independent check and safety arguments is that they seem to threaten *any* interesting version of the KK principle—even the qualified versions that allow for the possibility of knowing without believing that one knows. In the next subsection, I will present an example that puts pressure on the independent check and safety arguments.

Before getting into the details of my example, however, I want to emphasize some straightforward reasons to be suspicious of the independent check and safety arguments. The arguments are clearly onto *something*. The distinction between knowing, and knowing while also having an independent check on the reliability of one's source of information, is a real one and may very well be of some theoretical significance. Likewise with the distinction of being safe from error within a small sphere of nearby possibilities, and being safe from error within a larger sphere. But there are not strong pre-theoretical reasons to identify *these* distinctions with the distinction between knowing and knowing that one knows.

For one, it is not very plausible that in real conversations, questions about whether people know that they know are as easy to interpret as the anti-KK arguments would suggest; we do not hear questions about higher-order intrapersonal knowledge (as opposed to first-order knowledge) as straightforward questions concerning whether subjects satisfy extra high epistemic standards. It is no coincidence that when cataloging the various challenges facing the intelligence community, Donald Rumsfeld discussed the known knowns, the known unknowns, and the unknown unknowns, but failed to

---

[16] For example, Sydney Shoemaker, "Self-Knowledge and 'Inner Sense': Lecture II: The Broad Perceptual Model," *Philosophy and Phenomenological Research*, LIV, 2 (June 1994): 271–90, argues that believing that *P* and being fully rational entails believing that one believes that *P*. If we are sympathetic to this argument and sympathetic to KK, we might endorse a similarly qualified version of KK according to which it holds at least for all fully rational agents.

[17] In Greco, *op. cit.*, I offer a general "fragmentationist" strategy for responding to a number of putative counterexamples to principles like KK. While I do not explicitly discuss putative cases of knowledge without belief that one knows, I believe that the strategy I defend in that paper can be fruitfully applied to such putative examples.

mention the unknown knowns.[18] Had he included such a category, I doubt that people would have had a clear idea of what he had in mind.[19] Cases like that of the unwitting but reliable chicken-sexer (in which questions about higher-order intrapersonal knowledge seem to make sense) are *unusual*.[20] In real life, questions like "Sure she knows what time the movie starts, but does she know that she knows what time the movie starts?" would typically be met with puzzlement, rather than seamless accommodation. (If you do not believe me, try asking similar questions of non-philosophers.) While the defense of KK I will ultimately present is perfectly capable of drawing the distinctions drawn by the independent check and safety theorists, it refuses to identify these distinctions as distinctions between lower- and higher-order knowledge. This refusal not only has strong theoretical motivation—for example, in making room for common knowledge—but as the above considerations suggest, it fits with linguistic usage as well.[21]

I.3. An Example. Jones and Smith are playing a game of poker. The cards have been dealt and Jones has the stronger hand. Call the proposition that Jones has the stronger hand "*P*." As it turns out, the deal was videotaped by a reliable camera, and the tape clearly indicates who has which cards. Without looking at her cards Jones manages to procure a copy of the videotape, and she watches it.[22] Though Jones has no information about the reliability of the camera, she implicitly trusts it, and so comes to truly believe that *P*.[23] In this context, because coming to believe

---

[18] The exact quote is as follows: "[T]here are known knowns; there are things we know we know. We also know there are known unknowns; that is to say we know there are some things we do not know. But there are also unknown unknowns—the ones we don't know we don't know." U.S. Department of Defense News Briefing – Secretary Rumsfeld and Gen. Myers, February 12, 2002, 11:30 AM EDT.

[19] To be fair, even without including such a category, many commentators found his remarks to be less than a model of clarity.

[20] And as I argue in section III, they can be accommodated by the defender of KK.

[21] See also Berislav Marušić, "The Self-Knowledge Gambit," *Synthese*, CXC, 12 (August 2013): 1977–99, who stops short of defending KK, but does argue that it fits with key aspects of linguistic usage.

[22] I stipulate that she does not look at her cards because it will be important later that she has no independent checks on the reliability of the tape—if she looked at her cards, she would be able to compare what she sees directly with what appears on the tape.

[23] In realistic cases, subjects have *some* information about the reliability of video cameras—most of us know that they tend to work properly. However, for reasons that will come out towards the end of this section, I do not think that this unrealistic stipulation will affect the force of the example—I would still be able to make my argument using a version of the case in which Jones has a more realistic amount of information concerning the reliability of video cameras, though the argument would be slightly more complicated.

what one sees on the videotape is a reliable method for forming beliefs, it is plausible to say that Jones not only believes but *knows* that *P*.

The saga continues. Jones was taped while she was watching the tape. Smith sees this second tape—the tape of Jones watching the original tape. Like Jones, Smith implicitly trusts that the camera is reliable. So when Smith sees the tape of Jones watching the original tape, Smith comes to believe that Jones knows that she has the better hand.[24] If people generally do trust what is on videotapes (at least in contexts like this), and tapes really are reliable, then this seems like a reliable enough way of coming to believe that Jones knows that *P*; we should allow that Smith *knows* that Jones knows that *P*.

Let us go through one more iteration. Smith was taped when watching the tape of Jones. Jones sees the tape of Smith watching the tape of Jones watching the original tape, and so comes to believe (and by the same argument as above, know) that Smith knows that Jones knows that *P*. So at the end of all this tape watching, Jones knows that Smith knows that Jones knows that *P*.[25]

We have a three-stage story. At the end of the first stage, by watching a videotape, Jones knows that *P*. At the end of the second stage, by watching another videotape, Smith knows that Jones knows that *P*. At the end of the third stage, by watching yet another tape from the same camera, Jones knows that Smith knows that Jones knows that *P*. Write "$K_J P$" for "Jones knows that *P*" and "$K_S P$" for "Smith knows that *P*."

Let us now assume a closure principle on knowledge, to the effect that if *S* knows that *P*, and *P* entails *Q*, then *S* knows that *Q*.[26] At the end of the three-stage process, $K_J K_S K_J P$. Knowledge is factive, so if Smith knows that Jones knows that *P*, then Jones knows that *P*. So it follows from the content of what Jones knows—$K_S K_J P$—that Jones

---

[24] This may seem a bit too quick—in order for Smith to form this belief, she has to believe not just that the camera is reliable, but also that Jones believes this. Ultimately, this will not be important—we could set up the case so that both parties have excellent evidence that people always believe what they see on videotapes, or we might add that in each case in which somebody is taped watching a tape, he exhibits a sign of recognition which reliably indicates that he trusts what is on the tape.

[25] Perhaps all we can really say is that Jones knows that Smith *knew* that Jones *knew* that *P*. But we can set up the case to make it plausible that if Jones knows that Smith knew that Jones knew that *P*, then Jones knows that Smith knows that Jones knows that *P*.

[26] More realistic closure principles might not assume logical omniscience, and could be phrased in terms of being in a position to know that *Q* when one knows that *P* and *P* entails *Q*, rather than actually knowing that *Q* in such circumstances. This distinction will not matter in what follows.

knows that P—$K_J P$. So given our closure principle on knowledge, we get the result that $K_J K_J P$; Jones knows that Jones knows that P.[27,28]

In the remainder of this section, I will argue that our current conclusion—that Jones knows that Jones knows that P at the end of our three-stage story—is hard to reconcile with the independent check and safety arguments.

If we admit that Jones knows that Jones knows that P at the end of our story, then we have two options. Either (1) Jones knows that P (but not *that* she knows that P) after the first stage, and acquires new higher-order knowledge after the third stage, or (2) Jones already knows that she knows that P after the first stage of the story. I will argue that neither option is attractive for defenders of the independent check and safety arguments.

Why isn't option (1) attractive for the defender of the independent check argument? Because watching the second tape does not give Jones an independent check on the reliability of the camera. Of course, we might endorse (1) and hold onto the letter of the independent check argument by individuating belief-forming processes very finely. Instead of treating trusting the tape as a single belief-forming process, we might hold that for each tape, trusting what is on *that* tape constitutes a new, independent process. Then watching later tapes could provide independent checks on earlier tapes, and Jones could start out with only first-order knowledge,

---

[27] Saul Kripke, "On Two Paradoxes of Knowledge," in *Philosophical Troubles: Collected Papers, Volume 1* (New York: Oxford, 2011), also argues that considerations involving interpersonal knowledge are relevant to questions about the requirements for higher-order intrapersonal knowledge. Roughly, he argues that knowing that you know that P cannot be *that* hard, since it is often quite easy to know whether *other* people know that P, or for other people to know that you know that P.

[28] This might already seem odd, for a reason that the following modified version of the example brings out. Suppose Jones watches the original tape and then watches a tape of herself watching the original tape. It is very plausible that Jones has not acquired new higher-order knowledge by watching the tape of herself watching the original tape. To say that this *was* a way of gaining higher-order knowledge would be to endorse illegitimate bootstrapping, of the sort criticized by Vogel, "Reliabilism Leveled," this JOURNAL, XCVII, 11 (November 2000): 602–23; Adam Elga, "Reflection and Disagreement," *Noûs*, XLI, 3 (September 2007): 478–502; and others. If watching the tape does not provide new higher-order knowledge in the *intra*personal case, why should it do so in the *inter*personal case? While I have not yet introduced the materials necessary to fully answer this question, I can give a preview. If we accept the KK principle, we need not say that there is bootstrapping going on when Jones watches a tape of herself watching the original tape—watching the second tape will not be a way for Jones to *come to know* that she knows that P, since she will *already* have known that after watching the first tape. But accepting KK does not force us to say that once Jones watches the original tape, she also knows that *Smith* knows that P, or that Smith knows that *Jones* knows that P. So if KK holds, Jones's watching successive tapes can give her new higher-order *inter*personal knowledge, but not new higher-order *intra*personal knowledge, since there is no new higher-order *intra*personal knowledge to be had.

but acquire higher-order knowledge by watching another tape from the same camera (thereby getting an independent check on the content of the first tape). In the absence of a solution to the generality problem, we cannot rule out this strategy on principled grounds.[29] But it does seem to me that individuating processes at so fine a grain is contrary to the spirit of the independent check argument.

An example may help this point sink in. Suppose that you are in fact trustworthy, though I have no way of telling this. You tell me that *P*. The spirit of the independent check argument would suggest that I can know that *P* on the basis of your testimony, but cannot know that I (or you) know that *P*, since I lack an independent check concerning your trustworthiness. Suppose you go on to tell me that you are trustworthy—or, to make the case more precisely analogous to the example already under discussion, you tell me that somebody else knows that you know that *P*. It does not seem as if your making this further claim should let me know that I (or you) know that *P,* if I did not already know that you were trustworthy—if I do not know whether or not you are trustworthy, your telling me that you are cannot dispel my ignorance, at least *ceteris paribus*. The case involving the camera seems relevantly analogous—if I do not start out knowing that the camera is reliable, it is plausible that seeing more tapes from the camera in which the camera seems to be getting things right cannot dispel my ignorance. So, if the independent check requirement is to do the work it is meant to, such further tapes should not count as providing independent checks on the reliability of the camera.

Option (1) is just as unattractive for the safety theorist as it is for the defender of the independent check argument. There does not seem to be any sense in which watching the successive tapes is strengthening Jones's epistemic position with respect to *P* itself. The only information Jones has that bears on whether *P* is the original tape and other tapes of people watching that original tape. It does not seem as if possibilities in which the camera malfunctioned and misrecorded the players' hands are any more remote once Jones has watched tapes of people watching the original tape (or tapes of people watching tapes of people watching the original tape) than they were when Jones had only watched the original tape. While Jones seems to end up with higher-order knowledge concerning *P*, watching the second tape does not make her any safer with respect to *P* itself.

Perhaps the defenders of the independent check and safety arguments should pick option (2), and hold that once Jones has seen the

---

[29] See Earl Conee and Richard Feldman, "The Generality Problem for Reliabilism," *Philosophical Studies*, LXXXIX, 1 (January 1998): 1–29.

first tape, she not only knows but also knows that she knows that P. A defender of the independent check argument might hold that in any realistic case, Jones and Smith will start out with independent checks concerning the reliability of video cameras, and so Jones will be in a position to know that she knows that P already after watching the first tape. Along similar lines, a defender of the safety argument might hold that because video cameras are typically quite reliable, in any realistic case Jones would already be safe enough with respect to P to know that she knew that P—in realistic cases, there are not only no nearby possibilities in which cameras malfunction; there are also no nearby possibilities in which there are nearby possibilities in which cameras malfunction. This way the safety theorist could allow that Jones can come to know that Smith knows that Jones knows that P without becoming safer with respect to P.

While these moves are available, they do not go to the heart of the matter. Just as it is plausible that $K_J K_S K_J P$ holds at the end of the three-stage case, it is also plausible that after a five-stage case in which Smith and Jones each watched another tape, $K_J K_S K_J K_S K_J P$ would hold, or that after a seven-stage case in which they each watched yet another tape, $K_J K_S K_J K_S K_J K_S K_J P$ would hold.[30] If we accept these verdicts along with closure, then we must accept that in the original case, it is not just $K_J K_J P$ that holds at the first stage, but at least $K_J K_J K_J K_J P$. The guiding idea here is that Jones's acquiring higher-order interpersonal knowledge only requires that Jones acquire more information about *Smith's* epistemic situation—not that Jones acquire more information about what she herself knows. If we endorse the independent check or safety arguments, along with closure, we will have to reject that idea. We will have to accept that at some point (because Jones only has finitely many independent checks, or because there are *some* possibilities in which P is false, even if they are quite far away from the possibility Jones occupies), Jones cannot acquire another iteration of *inter*personal higher-order knowledge concerning what *Smith* knows by watching a tape of Smith, because Jones does not have enough iterations of higher-order *intra*personal knowledge.

Incidentally, this is why the independent check and safety arguments imply that common knowledge is unattainable; both arguments imply that in realistic cases, while we may often know that we know, our higher-order knowledge gives out at some point. Because we always have only finitely many independent checks, or because there are always possible cases in which we are in error (even if those cases are

---

[30] At least if Jones and Smith both have good enough memories to keep straight who has seen what, which we may assume they do.

only nearly nearly nearly nearly…nearby, for some large number of "nearly"s), we never know that we know that we know…*ad infinitum*.[31]

I hope I have made it plausible that the intuitive verdict about our example—that watching successive tapes is a way of acquiring higher-order interpersonal knowledge—does not sit nicely with the independent check and safety arguments, at least if we also accept a closure principle on knowledge. In the next section I will offer a simple formal model of knowledge that provides an attractive analysis of the case of Jones and Smith.

II. KNOWLEDGE, INFORMATION, AND NORMAL CONDITIONS

So far we have what looks like a puzzle. We have some *prima facie* attractive arguments against KK, along with an example in which they seem to prove too much. In the absence of some theoretical support for the example from the previous section, we might resolve this tension by simply dismissing the intuition that Jones knows that Smith knows that Jones knows that *P* after the third stage of this story, or at least dismissing the intuition that such a story could be extended to include arbitrarily many stages. My strategy in the remainder of this paper will be as follows. First, I will present a simple theory of knowledge on which KK and closure hold, and which I will argue provides an attractive analysis of the poker example from the previous section. The fact that our intuitive reactions to the example from the previous section can be accounted for in a simple, elegant model of knowledge militates against simply dismissing the intuitions. This will not yet resolve our puzzle—we will not yet have an explanation of what was attractive about the arguments against KK. I will go on to argue that when the model I am about to discuss is combined with contextualism concerning knowledge, we can explain away apparent failures of KK.

II.1. Normal Conditions and KK. Here is a simple, two-pronged analysis of knowledge: *S* knows that *P* just in case:

- *S* is in a state that carries the information that *P*, and
- *S*'s being in this state causes or constitutes *S*'s believing that *P*.

It is inspired primarily by Fred Dretske, but similar views have been defended by many other writers.[32] Moreover, as I argue below, even views that are not all that similar to Dretske's (for example, David

---

[31] For some reasons to think that this is a *bad* result—that is, that in many cases our mutual higher-order knowledge does not give out after some finite number of iterations—see Jane Heal, "Common Knowledge," *Philosophical Quarterly*, XXVIII, 111 (April 1978): 116–31, especially section III.

[32] See Fred Dretske, *Knowledge and the Flow of Information* (Cambridge: MIT, 1981). See also Stalnaker, *op. cit.*

Lewis's contextualist theory of knowledge) can be seen as sharing the same abstract structure, and it is this abstract structure that I will ultimately argue leads to KK.[33]

Of course, this theory is only helpful insofar as we have some idea of what it is to be in a state that carries the information that P. Here is how the phrase is intended to be understood: S is in a state X that carries the information that P just in case:

- Whenever conditions are normal, S is in X only if P, and
- Conditions are normal.

Analyzing the notion of information in terms of what would be true in certain conditions is a strategy that goes back to Dennis Stampe, and its many proponents include Dretske, Ruth Millikan, and Robert Stalnaker.[34] Of course, this further analysis is informative only insofar as we have some idea of what normal conditions are. It may help to start with an intuitive example. Suppose I see a dog. Conditions are normal in that my eyes are working properly, there are no misleading stimuli in my environment (for example, no papier-mâché dogs), and so on. In such a case, my brain (in particular, my visual system) will be in some state X such that, provided conditions are normal in the ways just gestured at, it will only be in X if there is in fact a dog in front of me. On the intended interpretation of "carries the information," in such a case my visual system carries the information that there is a dog in front of me.[35]

---

[33] Of course Dretske himself was no friend of KK. But just as contextualists hold that the Dretskean rejection of closure stems from a failure to appreciate the importance of context, I will suggest a similar diagnosis of the Dretskean rejection of KK.

[34] See Dennis W. Stampe, "Toward a Causal Theory of Linguistic Representation," *Midwest Studies in Philosophy*, II (September 1977): 42–63; Dretske, *op. cit.*; Ruth Millikan, *Language, Thought, and Other Biological Categories: New Foundations for Realism* (Cambridge: MIT, 1984); Millikan, "Biosemantics," this JOURNAL, LXXXVI, 6 (June 1989): 281–97; Stalnaker, *op. cit.*

[35] This way of putting things suggests that I am describing not normal conditions *simpliciter*, but normal conditions *for the operation of vision*. In fact, I think there are objections to the account in the text which may be best addressed by moving to a version of the normal conditions account that relativizes normal conditions to processes, such as vision. For instance, we might hold that S is in a state that carries the information that P just in case:

- S was caused to be in a state X by a process Y such that under all normal$_Y$ conditions, S is in X only if P.
- Conditions are normal$_Y$.

Here, the claim that conditions are normal$_Y$ is the claim that conditions are normal for the operation of process Y. The objections that would prompt moving to a process-relativized account will not be relevant to the arguments I will offer in the text, so I will stick with the simple account for ease of presentation. While it may not be obvious, the same arguments concerning KK that I will present in the context of the simple account would go through on the process-relativized account as well.

While some authors have tried to provide reductive, naturalistic analyses of normality—Stampe endorses such ambitions, but the most developed account I know of is provided by Millikan—we need not assume that a reductive analysis of normality is possible in order to be interested in the philosophical consequences of a normal conditions theory of knowledge. For instance, we might agree with Williamson and despair of finding a non-circular analysis of knowledge, while hoping that the above analysis, though in some sense circular (because we may need to appeal to the notion of knowledge in explicating our concept of normal conditions) can still lead to some interesting conclusions about the nature of knowledge.

Moreover, it is consistent with rejecting the hope for a *completely* reductive analysis that we try to say *some* informative things about normality. For instance, we might hold that which conditions count as normal is in some sense context-sensitive, and that our ascriptions of knowledge have to be understood as being made relative to potentially flexible background presuppositions about which conditions count as normal. Later in this paper I will invoke such a hypothesis to explain away some of the apparent problems with the KK thesis.

The class of "normal conditions" theories of knowledge, as I here understand that class, is a large one. This is because the sense of "normal" that I am using is quite thin. A theory counts as a normal conditions theory as long as it satisfies two conditions. First, it identifies some range of "normal" possibilities and says that you carry the information that P just in case you are in a state X that—as long as we restrict our attention to normal possibilities—you are only in if P. Second, it goes on to explain knowledge in terms of carrying information. But as I am using the phrase "normal," a theory counts as a normal conditions theory as long as it identifies some such set— it need not be the set of statistically normal possibilities, or biologically normal possibilities, or possibilities that are normal in any particular pre-theoretic sense of "normal."

In particular, pretty much any relevant alternatives theory of knowledge—that is, any theory according to which knowing that P requires being able to eliminate relevant alternatives to P, but where which alternatives count as relevant may vary from proposition to proposition, or context to context—will count as a normal conditions theory.[36] We can translate relevant alternatives theories into

___

[36] The idea of a "relevant alternatives" theory of knowledge comes from Dretske, "Epistemic Operators," this JOURNAL, LXVII, 24 (Dec. 24, 1970): 1007–23, but many epistemologists—especially contextualists of various stripes—have defended versions

normal conditions theories by holding that whenever an alternative is "relevant," the possibility in which that alternative is realized counts as "normal." To see how this works, it will help to consider an example.

Here is how David Lewis summarizes his contextualist relevant alternatives theory of knowledge:

> I say S knows that P iff P holds in every possibility left uneliminated by S's evidence—Psst!—except for those possibilities that we are properly ignoring.[37]

On Lewis's theory, one knows that P just in case one's evidence eliminates all of the relevant (that is, not properly ignored) not-P alternatives. Given how Lewis understands elimination by evidence, we can translate this into normal conditions talk as follows:

> S knows that P iff S is in an experiential state X such that in all normal (that is, not properly ignored) possibilities, if S is in X then P.

The translation makes it clear that Lewis's theory is a version of the normal conditions theory. His additions include the restriction to experiential states as the relevant type of information-carrying states, and a great deal of detail concerning which possibilities can and cannot be properly ignored (that is, which possibilities count as normal). His theory is contextualist (like many relevant alternatives theories) because which possibilities count as relevant/normal is determined in part by context.

We could use "normal conditions" more narrowly, such that a theory only counts as a normal conditions theory if it gives some particular sort of account of normality—perhaps statistical normality, or biological normality, or something along those lines. While the distinction between normal conditions theories in the narrow sense and the broader sense might be significant for some purposes, it is not for mine. In the remainder of this section, I will argue that normal conditions theories entail KK, and my argument will apply to any normal conditions theories even when we use the broad understanding of "normal conditions" theories according to

---

of relevant alternatives theories. Given that Dretske is the originator of both relevant alternatives theories of knowledge and normal conditions theories, perhaps it should not be surprising that we can see one type of theory as a special case of the other.

[37] Lewis, "Elusive Knowledge," in *Papers in Metaphysics and Epistemology* (New York: Cambridge, 1999), p. 436.

which pretty much any relevant alternatives theory will count as a normal conditions theory.[38]

To make good on this promise, however, I will need to make a simplifying assumption. In particular, in keeping with my focus on versions of the KK principle that abstract away from obstacles to KK involving failures of *belief* to iterate, I will assume that whenever agents are in states that carry the information that P, this causes or constitutes their believing that P.[39] Given this assumption, we can identify knowing that P with carrying the information that P, and we can ignore the belief requirement on knowledge. What, then, would it take for the KK principle to hold, on a normal conditions theory of knowledge? It would need to be the case that one's carrying the information that P entailed carrying the information that one carries the information that P.

I will first give an informal argument to the effect that in fact this *is* the case, and then I will provide a formal statement of the argument. Here is the basic idea. If S is in a state that carries the information that P, then *that very state* also carries the information that S is in a state that carries the information that P. Why is this? Higher-order information carrying requires that one be in a state that is correlated (given normal conditions) with being in a state that is correlated (given normal conditions) with P. But because every state is correlated with itself, if one is in a state X that is correlated with P, then one is also in a state (X itself) that is correlated with being in a state that is correlated with P.[40]

For example, suppose the current state of my stomach carries the information that I am hungry. My stomach is growling, and under normal conditions my stomach only growls when I am hungry.

---

[38] However, for reasons that will become clear, it will *not* apply to normal conditions theories that allow normal conditions to vary from proposition to proposition, in particular from P to $K_S P$. This is the view that Dretske himself endorsed, and it explains why he rejected KK and closure.

[39] This assumption is obviously false; I might be in a state that carries the information that my appendix has burst (for example, the state of having a burst appendix carries this information) without believing or knowing that my appendix has burst. But making this assumption amounts to abstracting away from the belief requirement on knowing. A defense of KK on this assumption would amount to the claim that—roughly—*if* the belief requirements for knowledge were always met, *then* KK would hold. Because the objections to KK I am concerned with in this paper turn on the idea that *even if* one has the necessary higher-order beliefs, one might still know while failing to know that one knows, this assumption is legitimate in the present context.

[40] Thanks to Bob Stalnaker for discussion here. See also Williamson, "Comments on Michael Williams' *Contextualism, Externalism and Epistemic Standards*," *Philosophical Studies*, CIII, 1 (March 2001): 25–33, who credits Lloyd Humberstone with pointing out (in conversation) that Lewis's theory vindicates some strong principles of epistemic logic, including KK.

If my stomach is growling, then my stomach is in a state (the growling state) that carries the information that it is in a state (the very same growling state) that carries the information that I am hungry. This is because whenever my stomach is growling, I am in a state—the growling stomach state—that carries the information that I am hungry. Higher-order information comes for free with first-order information carrying.

Here is the formal statement of the argument. Let N be the claim that conditions are normal, and let $\Box_N$ be a restricted necessity operator; $\Box_N P$ means that in all N-worlds, P holds.

By our definitions, S carries the information that P just in case:

(1) S is in a state X such that $\Box_N(S$ is in $X \supset P)$, and
(2) N.

What does it take for an agent S to have higher-order knowledge on the present account? S must be in a state that carries the information that the above two conditions (1 and 2) are met, and conditions must be normal. That is, if S is in a state that carries the information that S is in a state that carries the information that P, then:

(1′) S is in a state X' such that $\Box_N(S$ is in $X' \supset (S$ is in a state X such that $\Box_N(S$ is in $X \supset P)$ & $N))$ and
(2′) N.

To show that the present account of knowledge entails KK, it will suffice to assume that 1 and 2 hold and prove 1′ and 2′ from these assumptions. The proof is straightforward.

That 2′ follows from 2 is obvious—they are identical. Moreover, the last "N" in 1′ can be ignored—any claim of the form $\Box_P(Q \supset R$ & $P)$ is equivalent to $\Box_P(Q \supset R)$. So we are now faced with showing that 1″ (the same as 1′, but with an "N" removed) follows from 1:

(1) S is in a state X such that $\Box_N(S$ is in $X \supset P)$
(1″) S is in a state X' such that $\Box_N(S$ is in $X' \supset S$ is in a state X such that $\Box_N(S$ is in $X \supset P))$

If we let $X' = X$, we can see that 1″ follows from 1. We can rewrite it a third time, substituting X for X' in 1″:

(1‴) S is in a state X such that $\Box_N(S$ is in $X \supset S$ is in a state X such that $\Box_N(S$ is in $X \supset P))$

1‴, while hard to parse, is equivalent to 1. Our proof is now complete—on the present account, carrying the information that P (knowing that P) entails carrying the information that one carries the information that P (knowing that one knows that P).

The reasons why closure holds, in this model, are very similar to the reasons why KK does. Suppose *P* entails *Q*. Then if *S* is in a state that carries the information that *P*, *S* is in a state that carries the information that *Q*—if under normal conditions *S* is in *X* only if *P*, and *P* entails *Q*, then under normal conditions *S* is in *X* only if *Q*. In this model, knowing that *P* entails knowing that *Q*, if *Q* follows from *P*.

The example of the growling stomach may help again. Suppose that it follows from the fact that I am hungry that I am less than fully comfortable. If my growling stomach carries the information that I am hungry—that is, if under normal conditions, my stomach only growls when I am hungry—then it must carry the information that I am less than fully comfortable. If under normal conditions, my stomach only growls when I am hungry, then under normal conditions, my stomach only growls when I am less than fully comfortable.

This model has some counterintuitive features—KK and closure themselves can seem counterintuitive, and the model also implies that agents trivially know that conditions are normal whenever conditions are in fact normal.[41] Some counterintuitive consequences of the model can be deflected by pointing out that the simplifying assumption I made is false; often when one carries the information that *P*, one nevertheless does not *believe* that *P*, and so will not count as knowing that *P* according to the model I have been discussing. In the final section of the paper I will introduce another, general strategy for blunting counterintuitive consequences of the model— as a preview, we can hold that which conditions count as normal need not be the same from context to context, and that apparent failures of principles like KK and closure can be explained as cases in which focusing on certain questions induces a shift in context. Before introducing these complications, however, it will be fruitful to see how we might apply the current model—KK, closure, and all—to the example from the previous section.

In the poker example, it is natural to hold fixed what we mean by normal conditions—conditions are normal just in case the camera is functioning properly and it is commonly believed (or merely commonly implicitly trusted) that the camera is functioning properly.[42] As argued above, the present account of knowledge entails KK and

---

[41] Any state you could be in is a state such that, if conditions are normal and you are in that state, then conditions are normal.

[42] The definition of common belief is what you get when you substitute "belief" for "knowledge" in the definition of common knowledge mentioned at the beginning of this paper.

closure when normal conditions are held fixed. So the present account can explain why our initial description of the three-stage case was correct. Upon seeing the original tape, Jones comes to believe that she has the stronger hand, and because conditions are normal, and normally she will only have this camera-based belief if she *really does* have the stronger hand, she knows that she has the stronger hand. At the second stage, Smith acquires a camera-based belief that Jones knows that she has the stronger hand. Again, if conditions are normal—the camera is functioning properly, and this is commonly believed—then Smith will only have this higher-order camera-based belief if Jones really does know that she has the stronger hand. And so on at the third stage, and potentially beyond—the present account has it that seeing still further tapes would supply ever more iterations of interpersonal knowledge.

Because the account entails KK, it avoids positing obstacles to Jones's coming to know that Smith knows that Jones knows that *P* associated with obstacles to Jones's coming to know that she herself knows that *P*—because KK holds, knowing that one knows does not require some independent check or extra safety with respect to *P*. Because (intuitively) Jones *can* come to know that Smith knows that Jones knows that *P* just by watching another tape (and without having an independent check or getting any more safe with respect to *P* itself), this is an attractive feature of the present account.

We still, however, have not resolved our puzzle. Why is it that in some cases—like the case of Jones and Smith—it seems natural to think that higher-order knowledge does not require independent checks or extra safety, while in other cases—for example, the case of the reliable chicken-sexer—it seems natural to think that it does? In the next section I will offer a contextualist attempt at a reconciliation, along the lines of standard contextualist strategies for saving closure from apparent counterexamples.

III. KK, CLOSURE, AND CONTEXT

It is a standard part of contextualist lore that while closure holds *within* a context, it does not hold *across* contexts.[43] Hold fixed the context, and if I know that *P*, then I know that *Q*, if *P* entails *Q*. But it may be that once the question of whether *Q* is raised, I no

[43] Gail Stine, "Dretske on Knowing the Logical Consequences," this JOURNAL, LXVIII, 9 (May 6, 1971): 296–99, defends an early version of this strategy as a response to Dretske, "Epistemic Operators." For some other examples, see Stewart Cohen, "Knowledge, Context, and Social Standards," *Synthese*, LXXIII, 1 (October 1987): 3–26; Keith DeRose, "Solving the Skeptical Problem," *The Philosophical Review*, CIV, 1 (January 1995): 1–52; and Lewis, "Elusive Knowledge."

longer speak truly by self-ascribing "knowledge" of Q, since raising the question shifts the context, and thereby shifts the extension of "knowledge."[44]

For instance, consider Dretske's (*op. cit.*) example of a man and his son looking at a zebra at the zoo. Dretske rejected closure, and held that in such a situation, one might know that the animal in the enclosure is a zebra while at the same time failing to know that it is not a cleverly disguised mule. One motivation for this position comes from reflection on how it would be reasonable for the man to respond to various questions. For instance, in response to his son's question, "Do you know what kind of animal that is?," it would be quite reasonable for the man to respond, "Why yes, it's a zebra." But in response to the question, "How do you know that it is not a cleverly disguised mule?," it would be reasonable to respond, "I know no such thing; I am in no position to rule out that possibility." If we agree with these responses and take them at face value, as Dretske does, then it seems as if we must reject closure. After all, an animal's being a zebra entails its not being a cleverly disguised mule, so closure would imply that if one knows that an animal is a zebra, then one knows that it is not a cleverly disguised mule.

Contextualists have a different account of what is going on in cases like the one Dretske discusses. According to contextualists, in response to the initial question as to what kind of animal is in the enclosure, the man can speak truly by claiming to "know" that it is a zebra. In this context, before the question about cleverly disguised mules has been asked, the man "knows" not only that the animal is a zebra, but also that it is not a cleverly disguised mule; closure holds. But once the question concerning cleverly disguised mules has been asked, the extension of "knowledge" shifts, and the man speaks truly by denying that he "knows" that the animal is not a cleverly disguised mule. He would also speak truly by denying that he "knows" that the animal is a zebra—again, closure holds in the new context as well. Closure will not hold across contexts, however; what follows from what is "known" in the first context—for example, that the animal in the enclosure is not a cleverly disguised mule—need not be "known" in the second.

If we understand contexts as determining sets of normal conditions, then we have a nice explanation of this—as argued above,

---

[44] From here on out I will be sloppy about use and mention, and to avoid distracting the reader I will omit quotes in cases where, strictly speaking, they would be appropriate.

when normal conditions are held fixed, knowledge is closed under logical implication. When normal conditions vary, however, it need not be. *P* might hold in all normal conditions *N*, and *P* might entail *Q*, but raising the question of whether *Q* might bring us to a context with a new set of normal conditions *N′*, in some of which neither *P* nor *Q* holds.

Why opt for the contextualist account of the case rather than the simple one on which closure really does fail? One reason comes from the phenomenon of "abominable conjunctions."[45] While it can seem reasonable to say that the man knows he is looking at a zebra, and then to say that he does not know that he is not looking at a cleverly disguised mule, there seems to be something wrong with saying both in one breath. The following conjunction is abominable:

> (AC) The man does not know that the animal is not a non-zebra cleverly disguised to look like a zebra, but he does know that it is a zebra.

If we deny closure, it is hard to explain why anything should be wrong with sentences like (AC). But if we are contextualists, we can hold that while there are contexts in which "he knows it is a zebra" is true, and contexts in which "he does not know that it is not a non-zebra cleverly disguised to look like a zebra" is true, there is no context in which their conjunction, (AC), is true.

While diagnosing putative failures of closure along the above lines is by no means uncontroversial, strategies like it have seemed attractive to many contextualists. We might try to use a similar strategy, however, to diagnose putative failures of KK. Such a strategy would involve holding that within a context (that is, holding fixed normal conditions), KK holds. But in a context in which *S* knows that *P*, shifting our focus to the question, "Does *S* know that *S* knows that *P*?," might bring us into a new context in which *S* has neither second- nor first-order knowledge.[46]

---

[45] See DeRose, *op. cit.*

[46] In recent years rivals to contextualism have arisen that seek to explain much of the same data that contextualism is meant to explain. I have in mind subject-sensitive invariantism of the sort defended by Hawthorne, *Knowledge and Lotteries* (New York: Oxford, 2004), and Jason Stanley, *Knowledge and Practical Interests* (Oxford: Clarendon, 2005), along with relativism of the sort defended by John MacFarlane, "Relativism and Knowledge Attributions," in Sven Bernecker and Duncan Pritchard, eds., *The Routledge Companion to Epistemology* (New York: Routledge, 2011), pp. 536–44. It is not clear to me whether for my present purposes I need to take a stand on the issues that divide these newer theories from contextualist theories of the traditional sort. While I have focused on contextualist theories here, it may be that many or all of the same points could be made in a subject-sensitive invariantist or relativist framework. In particular, I suspect that where I appeal to context relativity, a relativist could appeal to assessor or standard relativity.

There are at least two questions we should ask before endorsing such a strategy. First, what is the *prima facie* motivation for pursuing it? Second, how will the details work—that is, why should asking questions about higher-order knowledge shift the conversational context in ways so as to make KK seem false?

III.1. Defending the Parallel. The *prima facie* motivations for pursuing a contextualist defense of KK are very similar to those that militate in favor of a contextualist defense of closure. Firstly, there are theoretical motivations. As I have already argued, normal conditions theories of knowledge entail closure and KK alike, when normal conditions are held fixed. If we defend closure by holding that apparent counterexamples involve context shifts, we should take a parallel line concerning KK.

But we can also appeal to the phenomenon of abominable conjunctions to motivate KK. Return to the case of the reliable chicken-sexer, whom we may name "Jimmy." Suppose Jimmy in fact has a reliable chicken-sexing faculty, and this faculty produces in him the belief that some chicken is female. Should we say that Jimmy knows that the chicken is female? In some contexts, this may seem perfectly reasonable. However, suppose we go on to urge that Jimmy does not know that he knows the sex of the chicken, because he cannot rule out the possibility that he has no reliable chicken-sexing faculty and is merely guessing. Once this is done, it is a bit awkward to insist that he nevertheless does know that the chicken is female—if Jimmy cannot rule out the possibility that he is just guessing, then how can he rule out the possibility that the chicken is not the sex he thinks it is? That is, how can he know what sex the chicken is? One way of bringing out the tension involves considering the following abominable conjunction:

> (AC′) Jimmy does not know whether he knows the sex of the chicken or is merely guessing, but he knows that he is not guessing incorrectly.

The opponent of KK must accept the first conjunct because she denies that Jimmy knows that he knows. Why must she grant the second conjunct? Suppose Jimmy knows that the chicken is female. It follows from the fact that the chicken is female, and from the fact that Jimmy believes that the chicken is female, that Jimmy is not guessing incorrectly. So if the opponent of KK grants that Jimmy knows that the chicken is female, and grants that Jimmy knows that he believes that the chicken is female, and grants closure, then she must grant that Jimmy knows that he is not guessing incorrectly.[47]

---

[47] If the opponent of KK rejects closure, she can also resist the conclusion that Jimmy knows that he is not guessing incorrectly. The argument that Jimmy knows

If we deny KK, then we will accept that one sometimes knows that *P* without knowing whether one knows or is merely guessing. But then we cannot explain why conjunctions like (AC′) are abominable—we cannot explain why there is a tension between acknowledging that for all Jimmy knows, he might be guessing (that is, he does not know that he knows the sex of the chicken), *and* insisting that he nevertheless does know the sex of the chicken. The contextualist theory, on the other hand, provides a satisfying explanation of how such a tension arises; there is no one assignment of normal conditions that makes true both the claim that Jimmy knows the sex of the chicken *and* the claim that he does not know that he knows. We can still explain why this does not seem obvious by holding that shifting our focus between the two questions tends to shift the sets of conditions that count as normal for the purpose of knowledge ascriptions. My strategy has been exactly analogous to arguments that have been offered on behalf of closure; just as denying closure leads to abominable conjunctions, so does denying KK.

Another parallel between extant contextualist defenses of closure and potential contextualist defenses of KK concerns the role of closure and KK in skeptical arguments. A common skeptical strategy is to appeal to some putative piece of knowledge that we lack—the knowledge that we are not being deceived by an evil demon, perhaps—and to then appeal to a closure principle to conclude that we lack other pieces of knowledge that we ordinarily take ourselves to have. Some anti-skeptics will respond to this style of argument by rejecting closure. Others will respond by insisting that we *do* know that we are not being deceived by an evil demon. But contextualists will offer a diagnosis much like the one I have already discussed concerning the case of the cleverly disguised mule—we can accept both skeptical claims to the effect that we lack certain pieces of knowledge, and ordinary commonsense claims to the effect that we have certain pieces of knowledge, *and* closure. According to the contextualist, the apparent conflict between commonsense knowledge attributions and skeptical denials of knowledge in fact stems from

---

that he is not guessing incorrectly implicitly relies on closure. If closure fails, then Jimmy might know that he believes that the chicken is male, and know that the chicken *is* male, but fail to know whether his belief about the maleness of the chicken is correct. I take it, however, that appealing to closure in the present context is dialectically appropriate. After all, I am arguing that there are underappreciated similarities between closure and KK, and that arguments for closure are mirrored by similar arguments for KK. My targets, then, are those who are already sympathetic to closure.

a failure to appreciate the fact that such attributions and denials are typically made in different contexts, in which different standards for knowledge are at play.

There is a parallel dialectic concerning KK and skepticism. Certain skeptical arguments start with the denial that we have some piece of higher-order knowledge. The skeptic might argue, for instance, that we do not know that we know that we have hands, perhaps on the grounds that such higher-order knowledge would require being able to meet certain challenges that we are in fact unable to meet. She can go on to argue that, because KK holds, if we lack higher-order knowledge, we also lack first-order knowledge—we do not know that we have hands. Adler (*op. cit.*) and Stroud (*op. cit.*) both consider arguments in this family, Stroud endorsing them, and Adler rejecting them on the grounds of opposition to KK. Contextualism suggests a natural strategy for avoiding full-blown skepticism, while still holding onto KK. Claims to the effect that we have first-order knowledge, and other claims to the effect that we *lack* higher-order knowledge, can both be true even if KK holds, so long as such claims are made in different contexts. The contextualist treatment of skeptical arguments that rely on KK, then, exactly parallels the contextualist treatment of skeptical arguments that rely on closure.

III.2. Context and Higher-Order Knowledge. Now that I have shown that the *prima facie* motivations for pursuing a contextualist defense of KK are exactly analogous to those that motivate pursuing a contextualist defense of closure, we can turn to the second question: why should asking questions about higher-order knowledge shift the context in ways so as to make KK seem false?

One simple explanation of why this might happen would appeal to salience. That is, we might hold that if some possibility is salient, then it should be included in normal conditions, and also hold that focusing on questions about higher-order intrapersonal knowledge typically makes new error possibilities salient.[48] While I think this is right as far as it goes, there is more that we can say.

If KK holds in all contexts, then asking whether somebody knows that she knows (after it is granted by all conversational participants that she knows) amounts to asking a question whose answer already follows from what is commonly accepted by the parties to the conversation. Since we generally try to interpret people so that the answers to their questions do not straightforwardly follow from

---

[48] See Lewis, "Elusive Knowledge."

what is already commonly accepted—plausibly, this is a special case of "accommodation" in David Lewis's sense[49]—we may interpret the question as implicitly inviting us to shift to a new context in which more possibilities count as normal.

Why would this make for a natural interpretation of the question? Even if we grant that somebody is in a state that is correlated with P in all "normal" possibilities N, it will typically be a non-trivial question whether she is in a state that is correlated with P in all possibilities N', where N' is a superset of N. So if we think that questions are typically interpreted so that their answers are not trivial consequences of what conversational participants already accept, and we are sympathetic to normal conditions accounts of knowledge, we should think that the question of whether somebody knows that she knows (asked in a context in which it is granted that she knows) will typically induce a shift into a new context where more possibilities count as normal, since this will be required if the question is to be interpreted as non-trivial.

For instance, in the case of the reliable chicken-sexer, when focusing on the first-order question of whether the chicken-sexer knows the sex of the chicken, we may presuppose a set of normal conditions that only includes cases in which her chicken-sexing faculty is reliable, and we will therefore grant that she knows. But suppose we now raise the question of whether she knows that she knows, and we try to suggest that it should get a negative answer—after all, by her lights, she might just be guessing! Such a shift in focus may bring us to a new context, where some conditions in which she does *not* have a reliable chicken-sexing faculty count as normal. It will not follow from the fact that the chicken-sexer counted as knowing the sex of the chicken in the old context that she will count as knowing that she knows in the new context—this is so even though, in the new context, there is no difference between knowing and knowing that one knows. For the chicken-sexer to be right in all the cases that count as normal in this new context (and thereby know, and know that she knows) might require, for example, having an independent check on the reliability of her chicken-sexing faculty. This would explain the appeal of the independent check argument—raising questions about higher-order intrapersonal knowledge may often bring us into a context in which higher-order intrapersonal knowledge (and first-order knowledge) requires independent checks.

[49] See Lewis, "Scorekeeping in a Language Game," *Journal of Philosophical Logic*, VIII, 1 (January 1979): 339–59.

Once we supplement our normal conditions theory of knowledge with contextualist machinery, we can also explain away the appeal of the safety argument. We must deny that safety is a requirement on knowledge if we endorse a normal conditions account of knowledge; the safety requirement is inconsistent with KK, and normal conditions accounts entail KK (holding context fixed). But we can explain why safety seems like a plausible requirement in the following manner. Suppose we try to find an intuitively plausible counterexample to safety. This would require identifying a possibility w, a subject S, and a proposition P such that it seems to us as if S knows that P in w, even though there is a nearby and relevantly similar possibility w' in which S falsely believes that P.[50]

But if explicitly considering a possibility tends to make that possibility salient, and thus induces a context shift so that w' counts as a "normal" possibility, then we should not expect to be able to find such a counterexample. Once we explicitly consider the possibility w', that possibility will count as normal. Because, in w, S is not in a state that allows her to rule out the relevantly similar possibility that she is in w', in the new context in which w' counts as a normal possibility, S will not count as knowing that P in w.

Counterexamples to the safety requirement, then, might be "elusive," in the sense described by Lewis. While in any given context there may be instances of the safety requirement that are false, explicitly considering those instances will typically bring us into a new context in which they are true. So it is unsurprising if the safety requirement seems true—its instances will typically be such that, once they are explicitly considered, they *are* true. Incidentally, if this strategy for explaining away the appeal of the safety requirement is successful, it has implications beyond the debate over KK. The safety requirement is a crucial premise in Williamson's

---

[50] In describing what would be required for a counterexample to safety here, I am a bit more faithful to Williamson's version of safety than I was earlier in the paper—I refer to "relevantly similar" nearby possibilities, and require that they be possibilities not just in which P is false, but in which the subject falsely believes that P. When we do not make these qualifications, finding counterexamples to safety is a good deal easier. See Ram Neta and Guy Rohrbaugh, "Luminosity and the Safety of Knowledge," *Pacific Philosophical Quarterly*, LXXXV, 4 (December 2004): 396–406. Even making these qualifications, there might be other plausible counterexamples to safety. See for example Brian Weatherson, "Luminous Margins," *Australasian Journal of Philosophy*, LXXXII, 3 (September 2004): 373–83; and Selim Berker, "Luminosity Regained," *Philosophers' Imprint*, VIII, 2 (March 2008): 1–22. The argument in the text can be seen as a strategy of last resort. While we might be able to find plausible counterexamples to safety, even if we cannot, we have good reason to think that counterexamples, if they existed, might be elusive.

enormously influential "anti-luminosity" argument, which he and others have put to use in a wide variety of philosophical contexts.[51]

An additional reason for favoring a contextualist defense of KK against the independent check and safety arguments is that it has the attractive feature of explaining why higher-order *inter*personal knowledge typically *does not* strike us as requiring independent checks on the reliability of our first-order belief-forming faculties, or extra safety concerning first-order propositions. Even if it is granted by all conversational participants that *S* knows that *P*, in a normal conditions account of knowledge, nothing follows concerning what *S* knows about what *other* people know. So in a context in which it is presupposed that Jones knows that *P*, asking whether Jones knows that Smith knows that *P* will typically *not* shift the context—the answer to this question need not trivially follow from what is already presupposed, so there is no pressure to shift to a new context in response to the question. If the appeal of the idea that higher-order knowledge requires independent checks or extra safety comes from context shifts, then we can explain why those ideas are not appealing in cases in which we are discussing *inter*personal higher-order knowledge; it is because there is no pressure to shift the context when we ask questions about higher-order interpersonal knowledge.

Incidentally, my discussion thus far has been quite concessive to the opponent of KK, in the following respect. As I mentioned at the end of section I.2, in most real life examples, it is *not* intuitive that higher-order intrapersonal knowledge requires independent checks or extra safety, because it is not at all clear how to make sense of questions about whether people know that they know, given that they know. Cases like that of the unwitting but reliable chicken-sexer are unusual precisely because there is a natural "fallback" set of normal conditions that we can move to when questions about higher-order knowledge are raised. My contextualist explanation of the "apparent" failure of KK applies only to such cases—cases in which it is clear how to move into a new context that will make questions about higher-order knowledge legitimate, even after having granted (in the old context) claims about first-order knowledge. In cases where there is no natural set of fallback normal conditions—most cases, I suspect—the contextualist strategy I have been exploring suggests that it should be hard to interpret questions about whether people know that they know; it should be hard to hear such questions as legitimate requests for information.

---

[51] See Williamson, *Knowledge and Its Limits*, chapter 4.

This prediction is borne out when we consider questions not about second-order intrapersonal knowledge, but about intrapersonal knowledge of even higher orders. As I have already suggested, it is not always easy to interpret questions like the following: "I grant that Jane knows what time the movie starts, but does she know that she knows what time the movie starts?" But while such questions are at least sometimes intelligible, it is much rarer that we can make sense of analogous questions concerning third- or fourth-order knowledge, let alone knowledge of even higher orders. Without bringing in heavy-duty philosophical theory, there is no natural way to interpret questions like the following: "I grant that Jane knows that she knows that she knows what time the movie starts, but does she know that she knows that she knows that she knows what time the movie starts?" This should be puzzling to advocates of the independent check or safety arguments; on those views, the distinction between first-order and second-order knowledge is of a kind with the distinction between second-order and third-order knowledge, or between third-order and fourth-order knowledge; in each case, achieving the next iteration of knowledge just requires more of the same (either more independent checks, or more safety). Such views have a tricky time explaining why it should be any harder to interpret questions about knowledge of very high orders than it is to interpret questions about second-order knowledge.

The arguments of the present paper, however, suggest a natural explanation. In many situations, there are at least two sets of standards, one stricter than the other, to which we might hold knowledge attributions (in the present framework, a stricter set of standards corresponds to a wider set of normal possibilities). We can exploit this fact to interpret questions that presuppose the failure of KK; we can make sense of such questions by associating higher-order knowledge attributions with stricter standards than those of first-order knowledge attributions. But there need not be an infinite hierarchy of ever stricter standards that could provide natural interpretations for questions about intrapersonal knowledge of arbitrarily high orders. If it were in the nature of knowledge itself that each iteration of higher-order knowledge were more difficult to acquire than the previous one, it is hard to see why such a hierarchy should not always be available to provide interpretations for such questions. But if apparent failures of KK stem from context shifts in the manner that I have suggested, then it is not so surprising that our ability to interpret questions about higher-order intrapersonal knowledge gives out relatively quickly—at least when we try to presuppose that higher-order knowledge does not come for free.

The reason is that in most situations, there are only finitely many natural choices for how strict a set of standards to apply to knowledge attributions—only finitely many "fallback contexts" to which we might move when forced to consider a question about knowledge of a still-higher order.

IV. CONCLUSIONS

In this paper I have offered both defensive and offensive strategies to KK sympathizers. When playing defense, a KK sympathizer can exploit the analogy to closure, and avoid counterexamples by adopting the position that KK holds within contexts but can fail across contexts. When playing offense, a KK sympathizer can argue that her theory allows her to explain various phenomena that are puzzling when we deny KK. First, she can account for the example from section I.3—she can explain why higher-order interpersonal knowledge does not seem to require independent checks or extra safety, even though higher-order intrapersonal knowledge *seems* (at least sometimes) to require meeting such further conditions. Second, she can straightforwardly explain why conjunctions like (AC′) are abominable by holding that such conjunctions are false in every context.

We can distinguish between easier-to-achieve and harder-to-achieve epistemic statuses—in the contextualist framework, this distinction is reflected in the fact that it is harder for a belief to count as knowledge in some contexts than in others. We can also distinguish between first-order knowledge and higher-order knowledge. Many contemporary epistemologists run these two distinctions together, holding that the latter distinction is a special case of the former; according to them, first-order knowledge is comparatively easy to get, second-order knowledge harder, third-order knowledge still harder, and so on all the way up. Sometimes—for example, when discussing unwitting but reliable chicken-sexers—running these distinctions together is very natural, and largely harmless. But other times—for example, when discussing higher-order interpersonal knowledge, and especially common knowledge—conflating these distinctions gets in the way of fruitful theoretical projects; if we hold that higher-order intrapersonal knowledge is harder to achieve than lower-order intrapersonal knowledge, we erect barriers to higher-order *inter*personal knowledge that we may not have intended to put in place. Once we see that one can recognize both distinctions without conflating them—the contextualist approach I have defended does just that—we should take seriously the possibility that KK might be OK.

DANIEL GRECO

Yale University