Intuitions and Experiments

Jennifer Nagel, University of Toronto

Contemporary epistemologists employ various methods in the course of articulating and defending their theories. A method that has attracted particular scrutiny in recent years involves the production of intuitive responses to particular cases: epistemologists describe a person making some judgment and then invite their audiences to check this judgment's epistemic status for themselves. "Does S *know* that P?"— A well-constructed case can elicit a powerful intuitive verdict. But the power of intuitive responses is somewhat mysterious, for reasons to be discussed in some detail in what follows. One crucial characteristic of intuition is that its workings are not exposed to us at the moment of judgment. When we are asked whether a subject in some scenario has knowledge, an answer may come to mind more or less forcefully, without its being immediately transparent to us exactly why the answer has the valence or the force that it does.

This failure of transparency doubtless makes it easier for us, as self-conscious epistemologists, to wonder about the epistemic legitimacy of the method of cases. Experimental philosophers have suggested that close attention to the mechanics of intuitive judgment should make us uneasy about the method: intuitions may exhibit worrisome instability either within an individual (Swain, Alexander, & Weinberg, 2008, 335), or between groups of individuals, perhaps along such epistemically scary fault lines as ethnicity (Weinberg, Nichols, & Stich, 2001) or gender (Buckwalter & Stich, 2011). If a certain intuitive response to some Gettier case works for the majority of epistemologists, we have no guarantee that it will work for others outside our professional club, and no good reason—the experimentalists argue—to suppose that it supplies genuine evidence about the nature of knowledge. On the experimentalist way of thinking, philosophers have taken false comfort in achieving consensus (or near-consensus) amongst themselves on particular cases, and need to be reminded of the difference between agreeing with one's peers about something and being right about it. According to one group of experimentalists, "experimental evidence seems to point to the unsuitability of intuitions to serve as evidence at all" (Alexander & Weinberg, 2007, 63). Whatever exactly is going on when philosophers solicit intuitions and use them as premises in their arguments, it's thought to be epistemically questionable, and rather different in kind from what goes on when people engage in a legitimate cognitive enterprise like empirical science.

I agree with the experimentalists that the question of the epistemic status of epistemic intuitions is an excellent question, and that we can take some steps towards answering it by taking a close look at the empirical facts about intuition. I also agree that philosophers risk mistaking consensus for correctness when weighing the evidential value of their intuitions. In fact, according to the best available model of the relationship between subjective confidence and accuracy of intuition, the strength of an intuition generally correlates with its consensuality rather than its correctness. However, this model applies equally to domains whose epistemic legitimacy the experimentalists would not want to dispute; in particular, it applies to perceptual judgments. Rather than undermining the case method in epistemology, close attention to the mechanics of intuitive judgment reveals some deep similarities between epistemic intuition and perception.[1]

Section 1 begins with a general discussion of intuitive judgment and then isolates the type of intuition that serves a significant dialectical role in contemporary epistemology. To be dialectically effective, an intuition about knowledge does not need to be correct: it just needs to be shared by one's audience. It is an interesting question how skilled practitioners of the method of cases can know in advance which cases will resonate with their audiences: Section 2 shows how that question would be answered within the leading model of confidence in intuitive judgment, Koriat's Self-Consistency Model (SCM). On this model, the stronger an intuition feels to an individual, the more likely it is to be shared with others. If the SCM applies to epistemic intuitions, variations in the strength of epistemic intuition cannot be read at face value: some intuitions may be strong and widely shared but inaccurate. However, the fact that we can dissociate the strength and accuracy of epistemic intuitions does not entail that these intuitions generally lack evidential value. Similar dissociations arise within perception: we call them perceptual illusions. As Descartes observes in the First Meditation, the fact that our natural perceptual capacities have some vulnerability to illusion does not oblige us to be generally skeptical about those capacities. And if our epistemic intuitions arise from a similarly reliable natural capacity, they could be similarly trustworthy.

Section 3 argues that pre-theoretical epistemic intuitions do arise from a generally reliable natural capacity. Known to psychologists as 'folk psychology' or 'mindreading', this is our ordinary

---

[1] The observation that intuition and perception have some epistemically important features in common has been made by various defenders of armchair epistemology (e.g. Chudnoff, 2010; Sosa, 2007; Williamson, 2007). The similarities I examine in this paper have not previously attracted much attention, however.

resource for ascribing states of knowledge, belief and desire. Experimentalists who challenge epistemic case intuitions do not reject all intuitive capacities—and cannot do so, on pain of collapsing into a general skepticism. At least one prominent experimentalist has explicitly identified mindreading as the kind of intuitive capacity that can be trusted, and not without reason. It is entirely plausible that this capacity is largely reliable in its deliverances, not least because our intuitive mindreading generates predictions about what others will do and say, and these predictions—including predictions about the differences between thinking and knowing[2]—are subject to feedback and correction over time. Our mindreading capacities are universal; this section also presents evidence that neither ethnicity nor gender has a significant impact on knowledge ascription in general, nor on epistemologically interesting cases in particular. In addition, the majority of untrained subjects produce the standard philosophical responses, further evidence that we are drawing on a common capacity in the case method.

Although perception and mental state attribution are both largely accurate natural capacities, both are subject to certain natural illusions. Where strength and accuracy of intuition come apart we have various resources available for correcting ourselves; Section 3 concludes by noting some similarities between the resources available for self-correction in the domains of epistemic intuition and perceptual judgment. For the purposes of this paper I assume that it is epistemically legitimate to take sensory observations as yielding evidence about the physical world; the aim is to show that we have roughly similar reasons to take epistemic intuitions as yielding evidence about knowledge.

1. Intuitive judgment in general, and dialectically useful epistemic intuitions in particular

There are many characterizations of the split between intuitive and non-intuitive judgments, both within philosophy and within psychology (for reviews, see Evans, 2007; Nagel, 2007; Sloman, 1996; Stanovich & West, 2000). Fortunately, there is very considerable common ground between the major theories of intuition, and for present purposes it matters only that epistemic intuitions count as intuitive in some relatively uncontroversial sense. It should be fairly unsurprising that epistemic intuitions are intuitive in some mainstream psychological sense of 'intuitive', but a brief review of a

---

[2] The identification of knowledge as a mental state is uncontroversial within the psychology of mental state ascription; this issue will be discussed in more detail below.

psychological account may help to make this evident, and help to justify the application of psychological theories of intuitive confidence to epistemic intuitions in particular.

Both in philosophy and in psychology, intuitive judgments are seen in contrast to the judgments we produce through deliberate reasoning. Because this point is emphasized in a particularly clear fashion by Hugo Mercier and Dan Sperber, I follow the outlines of their view in what follows, and use their terms 'intuitive' and 'reflective' for the two contrasting kinds of judgment. Mercier and Sperber describe intuitive judgments as generated by 'processes that take place inside individuals without being controlled by them' (Mercier & Sperber, 2009, 153). The spontaneous inferences produced by these processes modify or update what we believe 'without the individual's attending to what justifies this modification' (ibid.).

Some care is needed in handling the claim that we do not attend to the factors justifying our intuitively produced beliefs. Failure to attend is not a kind of blindness or insensitivity. When we read the emotions of others in their facial expressions—to take an example of an uncontroversial case of intuitive judgment—neurotypical adults are remarkably accurate at detecting and decoding the minute shifts in brow position and nostril contour that distinguish emotions such as surprise and fear (Ekman & Friesen, 1975). But judgments reflect these cues without *our* attending to the cues: the cross-culturally robust ability to recognize basic emotions does not depend on any personal-level attention to the facial configurations and movements that justify these swift intuitive classifications (Ekman, 1989; Ekman & Friesen, 1975). What we are explicitly aware of at the personal level is the emotion we see expressed in the face of another; we are not typically aware of relevant set of subtle facial cues as such, let alone the subpersonal processing that connects them with the conceptual template of the corresponding emotion. So we do not need to have an explicit theory of the facial differences between surprise and fear in order to discriminate these conditions successfully in the faces of others, and indeed we may have considerable difficulty in formulating an explicit theory that reflects our actual competence. In intuitive judgment we are 'conscious only of the result of the computation, not the process.' (Sloman, 1996).

In reflective judgment, by contrast, we engage in explicit reasoning and devote personal-level attention to the grounds of the conclusions we reach. Because of the strict capacity limitations on conscious attention, reflective thinking is sequential in character; where intuitive judgment can integrate large amounts of information very rapidly in associative parallel processing, reflective judgment is restricted by the bottleneck of limited working memory space (for detailed discussion, see Evans, 2007). However, what it lacks in speed, reflective judgment makes up in flexibility (a

point particularly emphasized in Stanovich, 2005). Intuitive judgment supplies routine answers, say, automatically and effortlessly recognizing the face of a friend, or producing an answer when one is asked to sum 2+2; reflective judgment can tackle novel problems, for example applying a controlled sequence of operations to sum a set of large numbers we have not previously encountered. The controlled sequential character of reflective processes like complex arithmetical calculation and conscious deductive reasoning keeps them open to view: having engaged in an explicit process such as syllogistic reasoning or long division it is not particularly mysterious, from the first-person perspective, why we have arrived at a given answer.

The contrast between the intuitive and the reflective does not mean that these two types of processing are isolated from one another. They are intimately connected in various ways. To begin, reflective processes take the products of intuitive processes as input: in the reflective calculation of a novel long division problem, for example, the individual steps consist in the application of intuitive single-digit calculation. The integration between these kinds of processing can also go the other way. We can also have intuitive responses to the results of reflective calculations; for example, we may have intuitively generated feelings of surprise or relief when explicit deduction produces an unanticipated or desired result.

Complex relationships between controlled and automatic processing may also make it less than obvious which kind of processing is engaged in some task. Because a sense of effort accompanies the controlled allocation of attention in reflective judgment, doing something like long division feels hard while merely adding 2+2 feels easy. But energy can be applied differently to different parts of a task. As Mercier and Sperber point out, a sense of effort can also accompany a personal-level decision to keep attending, over a period of time, to the output of some particular intuitive module, like the intuitive capacity for face recognition (Mercier & Sperber, 2009). It takes effort to sustain the task of scanning faces in a crowd, searching for a friend, although in each instance the processing that decides whether or not a given individual is recognized is itself subpersonal and effortless. It may also take effort to scan an epistemological scenario for the presence or absence of knowledge, even if the resultant recognition of knowledge or mere belief is itself processed intuitively.

The sense in which epistemic intuitions are intuitive is a delicate matter.[3] When we read a scenario, we are often aware of a series of considerations that could bear on whether the subject of the case has knowledge. We may be particularly conscious of this when reading a scenario that invites us to represent or retrace the steps of a subject who is thinking reflectively herself. I have argued elsewhere that our epistemic evaluations are sensitive to whether the judgments being made by the protagonist of the case would naturally be made intuitively or reflectively (Nagel, 2011). By default, and in routine circumstances, we think (and expect others to think) intuitively. The extra effort of reflective thinking is triggered (and anticipated to be triggered in others) by such factors as high stakes, novel or unusual cases and the need to negate hypothetical possibilities. Consequently we tend to have somewhat different expectations about a routine judgment, like the recognition that a certain zoo animal is a zebra, as contrasted with a more effortful judgment like the judgment that the same creature is not a cleverly disguised mule. The reader of a scenario needs to engage in reflective cognition to represent the second type of judgment, or to grasp high-stakes judgments, like the judgments of a person who is exceptionally anxious about whether the bank will be open tomorrow, and actively contemplating possibilities involving changes in the bank's hours. However, this is not to say that classifying the results of such reflective judgments as knowledge (or mere belief) would itself need to be a reflective matter. The fact that there are various differences between intuitive and reflective judgments does not entail that there are any differences in the qualities that make the products of either type of judgment count as knowledge.

As an example, if the reliability of an observed subject's (reflective or intuitive) belief-forming process is a quality that helps to determine whether the subject knows, then intuitive knowledge ascription could be sensitive to differences in perceived reliability without any explicit personal attention on our part to the issue of reliability as such. Even if we need some mixture of intuitive and reflective cognition to follow the story, it is possible that we use intuitive processing across the board in determining whether the key mental state in the story is an instance of knowledge or mere belief.

---

[3] This question is further complicated by the diversity of opinion on how to characterize intuition: there are certainly understandings of 'intuitive' on which not all epistemological case responses would count as intuitive. For example, the line between intuitive and non-intuitive judgment is sometimes drawn strictly in terms of executive function, and there is evidence that mental state attribution may under some conditions draw upon executive resources (Apperly, Back, Samson, & France, 2008; Apperly, Riggs, Simpson, Chiavarino, & Samson, 2006). The interpretation of these results is somewhat controversial, however (Cohen & German, 2009), as is the more general question of how exactly the intuitive/non-intuitive distinction could best be characterized (e.g. Evans, 2007). The problems here lie beyond the scope of the present paper; for present purposes, it is enough to observe that spontaneous verdicts on epistemic scenarios are not generated by the kind of fully transparent process of reasoning we see in long division or the application of an explicit theory.

We have a variety of evidence that our identification of knowledge would generally be intuitive. Most notably, we do not need to possess or apply any explicit theory of knowledge in order to gain the sense that the protagonist of some scenario has or lacks knowledge. We are not fully conscious of the grounds of our judgment in the way that we would be if we were making a reflective categorization on the basis of an explicit theory. In making up our minds about some particular case, we are not typically conscious of matching features of the subject's judgment to features of some theory of knowledge. It is not clear that such a theory is available to us in any event. Even the theorist who does adopt an explicit working theory of knowledge—say, someone attracted to Goldman's early causal theory (Goldman, 1967)—can readily find himself moved to classify novel cases in ways that directly conflict with this working theory (as in Goldman, 1976). The resilience of the Gettier problem suggests that it is difficult (if not impossible) to develop any explicit reductive theory of knowledge that fully captures our actual patterns of response to particular examples (Williamson, 2000; Zagzebski, 1994).

Further evidence for the intuitive character of knowledge recognition could be found in the speed and frequency of our real-time decisions between verbs of thinking and knowing—'knows' and 'thinks' are both heavily used, ranking at #8 and #12 of the Oxford English Corpus list of our most common verbs. Judgments naturally become intuitive when made very frequently. While critics of philosophical methods have focused on deliberate assessments of hypothetical scenarios, we have reason to believe that the abilities brought to bear on those cases would be equally operative in our very frequent and spontaneous assessments of real-life situations (Saxe, 2006; Williamson, 2007). For example, similar brain regions are activated when experimental subjects read narratives about the knowledge and beliefs of others (but not about their appearance or subjective states such as hunger or thirst), and when they are engaged in live interaction with others, playing a game that requires them to attribute knowledge or belief to a partner (Redcay et al., 2010; Saxe & Wexler, 2005). Just as the intuitive mechanisms enabling recognition of emotion are stimulated in similar ways by live faces, videotapes of actual faces or animated cartoons, our resources for rapidly detecting knowledge and belief are thought to be stimulated in similar ways by live interactions and appropriate narrative representations.[4] Where attributions of emotion are

---

[4] Presentation format may however affect the degree to which the key resources are activated. In a recent fMRI study of responses to closely matched animated and live-action movie sequences, Raymond Mar and collaborators found just the same mental state attribution areas activated, but to a greater degree by the live-action sequences than by their cartoon counterparts (Mar, Kelley, Heatherton, & Macrae, 2007). Activation of these regions was spontaneous; participants were instructed to watch the movie clips closely but were given no instructions to make mental state inferences concerning the characters (cf. Cohen & German, 2009).

triggered by certain patterns of facial features, attributions of mental states are triggered by the recognition of patterns of 'input' to another's perceptual and inferential capacities—for example, by our automatic calculations of what an observed agent can see (Samson, Apperly, Braithwaite, Andrews, & Scott, 2010)—and by recognition of an agent's 'output' of intentional action and speech (for reviews, see Apperly, 2011; Goldman, 2006).

To suggest that our classifications of knowledge would ordinarily be intuitive is not to say that it is impossible to devise a situation in which we would think reflectively in identifying a judgment as an instance of knowing. For example, reflective classifications can be made when we are considering cases at a certain level of abstraction. Suppose we stipulate that there is some set of some properties Γ such that any judgment will instantiate knowledge if and only if it has all the properties in Γ, and then stipulate that judgment J has all the properties in Γ; the subsequent conclusion that judgment J is an instance of knowledge can be made reflectively, with full consciousness of the grounds of our categorization. Such formally acceptable but materially uninformative classifications are not what figure in the richer and more unpredictable case method that experimental philosophers have been attacking.

It should also be acknowledged that not all intuitive identifications of knowledge are equally rich and unpredictable. It should be possible to have theory-driven epistemic intuitions, for example, after becoming very well-rehearsed in applying the verdicts of some particular analysis of knowledge. Such classifications may fail to resonate with those who lack a commitment to that analysis, and would supply no independent evidence for the theory that produces them. As rote exercises in the application of an existing analysis, they would lack intuition's ordinary power to convince others, or to surprise us and reshape our theories. Some epistemic intuitions may indeed be theory-driven, but not all intuitions in current epistemology could have this status. The Gettier result that justified true belief was insufficient for knowledge did come as a surprise, and resonated with an audience of philosophers who had largely been committed to one or another form of JTB theory. At least some dialectically powerful epistemic intuitions are pre-theoretical.

It is an interesting question how the person who experiences a pre-theoretical intuition could see its value, especially when it runs against established theory. When Edmund Gettier found himself inclined to judge that *Smith does not know that the man who will get the job has ten coins in his pocket*, he could have dismissed this inclination as a mistake on his part. Given the broad acceptance of the JTB theory, it might seem Gettier could reasonably have concluded that he was subject to some momentary failure of insight or had some individual peculiarity driving him to

misclassify the case. He claimed instead that his verdict was clearly right, as if anticipating that he himself would make the same judgment about this case on other occasions, and expecting his audience to do likewise, notwithstanding their known prior inclinations to accept the JTB theory (Gettier, 1963). Gettier does not present an explicit argument showing exactly why Smith's judgment does not amount to knowledge: he does not offer any positive analysis of knowledge of his own, nor does he specify any necessary conditions on knowledge which are lacking in this case. But despite neither explaining nor perhaps even knowing exactly why he feels the way he does about this case, Gettier seems convinced that his intuition about it will be felt by others and by himself on other occasions. What is it about the character of intuitive judgment that could have made him feel that way? The next section examines this question.

2. Confidence and consensuality in intuitive judgment: Koriat's SCM

Although we are not conscious of the inner workings of intuition at the moment of judgment, we are conscious of certain differences in its deliverances: some intuitions feel clear and strong; others are weaker and more obscure. These differences in confidence have more than momentary and private significance. When an intuition is strong, it is likely to be stable, and likely to be felt by others. To a first approximation, this is because confidence in intuitive judgment is determined by the ease with which one makes the judgment (Alter & Oppenheimer, 2009; Kelley, 1993; Reber & Schwarz, 1999). In general, what is intuitively easy on one occasion is likely to be easy on subsequent occasions, and similarly easy for other people. An elegant model explaining this phenomenon has been developed in recent years by Asher Koriat. Using intuitive responses to two-answer forced-choice questions from a variety of domains, Koriat established that a person's level of confidence in an intuitive judgment predicts the degree to which that individual will make the same judgment again when presented with the same problem again, and, furthermore, predicts the judgment's consensuality—the extent to which others will make the same judgment (Koriat, 2008; Koriat, 2011).

Koriat's model of intuitive confidence, the Self-Consistency Model (SCM), traces its origins to some curious findings of his from the mid-1970s. In the course of investigating our ability to monitor our own memory performance, he decided to examine an obscure memory domain, where subjects would presumably not have had previous experience in monitoring their track record. According to the (now derelict) theory of universal phonetic symbolism, there are certain sounds which have robust cross-cultural significance. Certain sound-meaning relationships were

hypothesized to be encoded in the memory of all human beings. Some empirical evidence had been offered in support of the theory: for example, people perform at significantly better than chance when asked to match antonyms of value-laden, magnitude or sensation terms (beautiful, ugly; large, small; hard, soft) from their own language with those of a non-cognate language to which they have had no previous exposure (Slobin, 1968).

In Koriat's first study in this area (Koriat, 1975), he asked English-speaking participants to match 56 antonym pairs in Thai, Kannada and Yoruba to their English translations, presented in random order. After taking their best guess as to, say, which of the pair [tuun, luk] meant "deep" and which meant "shallow", participants were asked to rate their confidence—or what Koriat described in his instructions to them as "the unexplainable feeling that you may be correct"—in the choice they had made. Ratings were made on a 1-4 scale, where 1 was to be used for "a totally wild guess", and 4 for the answers about which most confidence was felt. On the matching task, Koriat replicated the pattern found by Slobin, with 21 word pairs yielding a significantly better than chance matching, and only 7 pairs yielding a significantly worse than chance performance. 49 out of his 55 subjects matched more than half the pairs correctly, the remaining 6 close to 50-50. The results of the confidence task seemed to suggest that participants had some access to the validity of their answers: only 53.35% of the pairs to which they had assigned confidence level 1 were correctly matched, versus 66.10% of the level-4 rated pairs. Accuracy rose monotonically with confidence rating across the board.

In a subsequent study, Koriat took a closer look at the relationship between confidence, accuracy, and consensuality, or the extent to which an intuition was shared across participants (Koriat, 1976). Drawing on earlier work, Koriat compiled a list of 85 antonym pairs in six non-cognate languages to which his participants claimed no previous exposure; each of these pairs was ultimately classified as either consensually correct (CC), consensually wrong (CW) or nonconsen-sual (NC). An antonym pair counted as CC if a statistically significant percentage of subjects (in this case at least 60%) selected the correct translation, as CW if a similar percentage agreed on the incorrect response, and as NC if neither response was significantly preferred. The 1975 study had shown a positive overall correlation between confidence and accuracy, but in that study CC pairs outnumbered CW pairs three to one, leaving it unclear whether confidence was really tracking correctness or consensuality. The new study analyzed the consensuality and reported confidence for each category of item. Confidence still correlated positively with accuracy for CC pairs, such as the Chinese [ching, chung], for which 94% of participants correctly chose [light, heavy] rather than

10

[heavy, light]. Across responses to items in the CC category, lower confidence ratings (1-2) were 69.63% likely to be accurate and higher confidence ratings (3-4) were 79.55% likely to be accurate. For nonconsensual pairs, there was no significant relationship between confidence and accuracy. Meanwhile, for CW pairs, such as the Hindi [ranjida, khush], for which only 18% of participants correctly chose [sad, happy], confidence and accuracy were negatively correlated: 34.45% of the low confidence answers were correct, compared with 24.22% of the high confidence answers. Strikingly, those who had selected a minority response (either the wrong translation for a CC item or the correct one for a CW item) tended to feel less confident about their choice than those who had chosen as the majority did. Koriat concluded that each individual's confidence was somehow "attuned to the consensuality of the response, regardless of its accuracy" (1976, 247).

One might wonder how isolated participants completing the task individually were able to respond as if tracking how others were responding; one might also wonder whether these findings are particular to the intuitive domain Koriat was probing, or whether they arise as a result of some general or structural features of intuitive judgment. Some progress on these questions has been made in recent years. There is evidence that underlying structural features of intuitive judgment are at work here. Results like those found in the phonetic symbolism task have been found for a wide variety of other two-alternative forced-choice tasks, including long-term memory for trivia (Koriat, 1995), short-term memory for sentences with and without interference from schema-based inference (Brewer & Sampaio, 2006), social attitudes (Koriat & Adiv, in press), and perceptual judgments (Koriat, 2011).[5] In the course of this work, two further correlations were uncovered: a participant's reported confidence level in any given judgment predicts not only the extent to which others will respond similarly, but also the likelihood that this particular participant will repeat that response when presented with the same two-alternative, forced-choice problem again later on (Koriat, 2008; Koriat, 2011). Furthermore, in a study which asked subjects to answer the same 50 intuitive questions seven times over the course of several days, a subject who fluctuated between different answers to an intuitive probe was more confident when he gave his majority answer (the answer that he gave more often) than when he gave his minority answer (Koriat & Adiv, in press).

These last correlations provided the guiding idea of the Self-Consistency Model (Koriat, 2011; Koriat & Adiv, in press). According to the SCM, subjective confidence is a byproduct of the

[5] Such results are not restricted to self-report of confidence, but have also been found for non-verbal behavior. For example, independent of their accuracy, more consensual answers attract higher wagers in betting tasks, even for subjects who have not been asked to give any explicit report of their level of confidence (Koriat, 2011; Simmons & Nelson, 2006).

process of intuitive judgment. The parallel processing that underpins intuitive and perceptual judgment draws on a vast range of information potentially relevant to the problem at hand. We do not consult the entire range on any given occasion. In generating an answer to a two-alternative forced-choice question, a sample of representations is drawn from the pool in order produce a response, where each representation is some consideration in favor of one or the other answer. Sampling continues until either a critical number of representations favoring one side has been amassed, or until a preset number of samples has been drawn.[6] One's response is determined by the direction of the sample's majority, and one's confidence in that response is determined by the size of the sample's majority, the internal consistency of the sample in one direction or the other. Where an almost equal number of representations in the sample speak in favor of each side, one's confidence will be much lower than it is when the sample is univocal. On this model, neither one's answer nor one's confidence level will directly measure the extent to which the underlying pool of information supports an answer to the question: it is possible to draw an unrepresentative sample from a pool which overall strongly favors P, and respond that not-P. However, when we have done such a thing, the odds are overwhelming that our sample will speak only weakly in favor of the unrepresentative choice.[7] If the pool strongly favors P, we are likely to draw a very consistent sample from it, and respond with higher confidence for P. If confidence is determined by the consistency of one's sample, then confidence will predict the likelihood of one's making the same choice on subsequent occasions. Furthermore, if others are drawing from a similar pool of representations, individual confidence will predict consensuality as well: a choice very heavily favored by one's own sample is likely to be favored by others also.

The SCM is a general model of intuitive judgment for two-alternative forced-choice questions. It is conceivable that intuitive judgments in epistemology are managed in some quite different way, although we have no positive reason to think so. The question asked at the end of the typical case—"Does S know that p?" is a two-alternative forced-choice question. The only published work bearing directly on the relationship between intuitive confidence and consensuality in

---

[6] While the focus of the present paper is intuitive judgment, it should be noted that similar sampling models have been proposed for reflective judgment. Benjamin Newell and Michael Lee, for example, have argued that we always sample evidence sequentially until we reach a preset threshold (Newell & Lee, 2010). Such models could extend the conclusions of this section to those who are inclined to think that epistemological case responses may to some extent be generated by sequential or reflective thinking.

[7] Koriat supplies the following example (using the abbreviation p_{maj} for the proportion of representations supporting the majority answer): 'with p_{maj} = .75, a sample of seven representations has a .445 likelihood of yielding six or seven representations that favor the majority choice. In contrast, the likelihood that it will yield six or seven representations that favor the minority choice is only .001' (Koriat, 2011, 121).

responses to epistemic scenarios has generated results directly in line with the predictions of the SCM (Wright, 2010).

If the SCM applies to intuitive judgments in epistemology as it applies to other types of intuitive judgments, the strength of an epistemic intuition is an index of its reproducibility. Just as a good experiment is one which can be reproduced by others or by oneself on other occasions, so also a good intuition. As long as the underlying pool of representations is similar for oneself and others, one's confidence in an intuitive response is a marker of its dialectical value: a strong intuition will typically be stable, and others will feel it, too. If Gettier felt strongly that his intuition would be shared by his audience, he was not wrong about that: whatever intuitive techniques he applied in categorizing the case—or whatever set of representations he pulled up from the pool to identify his Smith as failing to know—contemporary epistemologists have generally shared his experience. Even epistemologists who express reluctance to reject the JTB theory on the basis of Gettier intuitions do not deny the force of those intuitions (e.g. Weatherson, 2003). The applicability of the SCM to epistemic intuitions would help to explain the dialectical success of appeal to intuitions. If the SCM applies, however, there is a distinction between dialectical value and accuracy. Intuitive confidence is not directly correlated with the truth of one's response: there are some intuitive problems for which one's underlying pool of representations will have a robust tendency to generate the wrong answer.

On the side of perception, we classify these 'robustly and consensually wrong' cases as illusions. Our intuitive responses to classic illusions such as the Müller-Lyer can be clouded by our explicit knowledge of their deceptiveness. It is useful to consider some fresh cases, such as the following, taken from (Koriat, 2011):

Which of these line segments is longer?
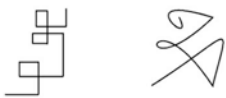
Which of these figures is larger?

In both cases, the correct answer is: the one on the right. However, the consensual answer for both cases is: the one on the left (endorsed by 84.62% of Koriat's participants in the first case and 82.93% in the second). These pairs of stimuli could be compared with the following, also from (Koriat, 2011):

Which of these line segments is longer?

Which of these figures is larger?

These last two pairs fall into the consensually correct (CC) category: a strong majority of subjects correctly identify the right-hand figure as larger and longer (83.59% and 89.75% respectively). The subjective experience of comparing the magnitudes is similar for deceptive and non-deceptive items of equal consensuality: we scan the figures until one answer or the other seems right. It is not transparent to us just what techniques we are applying in the course of making our evaluation. But whatever our techniques might be, they are sufficiently similar across persons that they show clear trends of working more swiftly and easily in some cases than in others. Whether or not they were being judged correctly, pairs on which there was more consensus were judged significantly more quickly and with significantly greater confidence than non-consensual pairs (Koriat, 2011).

Susceptibility to a common set of illusions is strong evidence of systematic similarities in our ways of judging magnitude. Ordinary and illusory cases are not distinguished in our immediate subjective experience, and common mechanisms are taken to underpin both; indeed, as Hermann von Helmholtz observed more than a century ago, perceptual illusions are "particularly instructive for discovering the laws of the processes by which normal perception originates" (von Helmholtz, 1893, 75). The fact that we are prone to certain illusions of the type illustrated above is of course not a reason for general skepticism about our capacity to distinguish magnitudes visually; rather, our general trust in this capacity continues to be warranted as long as the capacity is generally

14

accurate. Illusions are rare enough that there is overall a strong positive correlation between our confidence in our perceptual judgments and the accuracy of those judgments. Even setting aside illusions, we are not perfectly calibrated—we seem[8] to be somewhat underconfident about very easy judgments and overconfident about difficult ones (e.g. Baranski & Petrusic, 1994)—but in general, the perceptual judgments about which we feel most confident are substantially more likely to be right than the judgments about which we feel least confident. Similar patterns apply to other domains, such as long-term memory (Baranski & Petrusic, 1995). Where intuitive judgment is easy and swift for some intuitive capacity, we feel most confident; the slow and difficult judgments that arise at the borders of an intuitive capacity naturally generate less confidence (Alter & Oppenheimer, 2009). It is unsurprising that variations in subjective confidence would generally be adaptive in this manner: weak intuitive responses can signal us to be cautious or to supplement our intuitive responses with reflecting thinking in the interests of increasing accuracy (Alter, Oppenheimer, Epley, & Eyre, 2007). Natural intuitive capacities produce variations in subjective confidence that are meaningful and largely helpful.

In criticizing the case method, Jonathan Weinberg has suggested that intuition is "basically a 1-bit signal: is p possible, yes or no? Or: Does the hypothetical situation fall under the concept or not?" (Weinberg, 2007, 335) He acknowledges that we sometimes experience variations in the apparent force of intuition, but says that it is "completely unknown to what extent we have any inter- or even intra-subjective agreement about it." (ibid.) I have aimed to show that if epistemic intuitions work like other types of intuitive judgment, significant inter- and intra-subjective agreement is to be expected. Intuition is a signal which carries information not only about the yes-or-no question, but also about the extent to which one's own response will be stable and others will respond similarly. The way variations in subjective confidence are shared helps explain the dialectical success of the case method.[9] The fact that philosophers largely agree with each other

---

[8] Failures of calibration are controversial; Gigerenzer and collaborators have argued that an ecologically representative sampling of tasks would produce something closer to perfect calibration (Gigerenzer, Hoffrage, & Kleinbolting, 1991). Cesarini and collaborators argue that the 'hard-easy' effect is at least in part a statistical artifact, because error near the poles of the confidence scale can push us only towards the middle (Cesarini, Sandewall, & Johannesson, 2006).

[9] The experimentalist challenge to which I am responding here admits the dialectical effectiveness of the method of cases. Another possible challenge to the case method would dispute even that, perhaps suggesting that most philosophers do not actually feel the Gettier intuition but only pretend to. While it is true that there are some cases which provoke a divided response—and more will be said about such cases in Section 3—the suggestion that cases generally lack dialectical value is extremely hard to square with empirical facts about the practice of contemporary epistemology. In the debates between contextualists, relativists, interest-relative and strict invariantists, for example, there is almost uniform agreement on the intuitive power of the

about Gettier's cases is evidence that we are drawing from similar pools of representations in responding to them. However, the experimentalist can still wonder whether epistemologists' systematically shared intuitions are generally indicative of the nature of knowledge. Because intuitive confidence correlates with consensus rather than accuracy, the case method could be pursued with great dialectical effectiveness even if the intuitions on which professionals strongly tended to agree were typically shared illusions rather than shared insights. It is natural to wonder whether epistemic case responses arise from the sort of natural competence for which strongly felt intuitions are typically more likely to be right and illusions relatively rare, or whether the common pool from which epistemologists are drawing their shared intuitive judgments is an unfortunate product of training or selection effects. The next section tackles that issue.

3. Intuitive mindreading, amateur and professional

How do we attribute knowledge or mere belief to the protagonist of an epistemological scenario? The default answer to this question is: the same way we generally attribute knowledge or belief to anyone else. Unless there is a special reason to think that knowledge attributions work quite differently when we are reading philosophy papers—and I'll shortly survey some evidence against that sort of exceptionalism—we should expect to find that epistemic case intuitions are generated by the natural capacity responsible for our everyday attributions of states of knowledge, belief and desire. This capacity has been given various labels, including 'folk psychology', 'mindreading', and 'theory of mind'. I will follow Ian Apperly and others in calling this capacity 'mindreading', but intend to remain as neutral as possible about the current debates within psychology about the exact nature of this capacity.

Although a great deal of work in mindreading has focused on natural illusions of mental state attribution, particularly in children whose capacities are still immature, it is generally agreed that adult capacities for mental state ascription provide fairly reliable tracking of what others think and know. Interestingly enough, in his criticism of the case method, Weinberg explicitly exempts mindreading from the trouble zone of suspect intuitions: he is keen to establish that some intuitions are acceptable, in order to show that the experimentalist attack on philosophical intuitions does not

---

central cases, even where these cases are awkward or difficult for advocates of one or the other theory and it might be very tempting to claim one did not feel the relevant intuition (see e.g. DeRose, 2009; Stanley, 2005; Williamson, 2005).

have to collapse into a general skepticism. Weinberg very reasonably observes that mindreading intuitions "make all sorts of predictions about the world," and further notes that "we seem at least somewhat capable of learning from the occasional failure of those predictions" (Weinberg, 2007, 339). Many researchers in mindreading would make stronger claims about the ways in which our capacity to ascribe states of desire, knowledge and belief get shaped by feedback from successful and unsuccessful predictions, but Weinberg's minimal concessions are enough for present purposes. If Gettier's intuitions about what Smith does or doesn't know come from our everyday mindreading capacity for ascribing states of knowledge and belief, and if this capacity is generally reliable, then our epistemic case intuitions have some positive claim to epistemic legitimacy.

Because Weinberg himself explicitly locates Gettier intuitions in the trouble zone of suspect intuitions (2007, 335), he can't be understanding them as ordinary mindreading intuitions. This is probably not a capricious move on his part: his own experimental results could seem to make it very reasonable to take Gettier intuitions to arise from some capacity other than mindreading. Mindreading capacities are generally thought to be cross-culturally universal, not least because of developmental similarities in mindreading in radically different cultures (Wellman, Cross, & Watson, 2001). Natural illusions within the mindreading capacity are also cross-culturally shared.[10] Furthermore, mindreading capacities are thought to exhibit no gender differences in typical adults; extensive research on mental state ascription has shown only a mild advantage for girls before the age of four, perhaps simply as a function of girls' earlier linguistic development, and no discernable gender differences in the non-clinical population beyond that point (Charman, Ruffman, & Clements, 2002). For those who suspect that there is cross-cultural variation in Gettier case recognition (following Weinberg et al., 2001), and for those who suspect that Gettier intuitions may not be felt equally by men and women (following Buckwalter & Stich, 2011), it may seem unlikely that epistemologists' responses to Gettier cases would be produced by our common mindreading capacity.

Another barrier to allocating Gettier case responses to our common mindreading capacity could be the widespread philosophical opinion that belief is a mental state while knowledge is not.

---

[10] In particular, we share a bias known as hindsight or more broadly epistemic egocentrism, a bias which distorts our mindreading of those who occupy a more naïve perspective, including our own more naïve past selves. There had been some suggestions that this bias differs across cultures: for example, one study found that the bias is worse under certain instructions for Western (Canadian) subjects than for Eastern (Japanese) subjects (Heine & Lehman, 1996), while another found greater hindsight on the Eastern (Korean) than on the Western (American) side (Choi & Nisbett, 2000). More comprehensive work has undermined claims of cross-cultural differences in either direction (Pohl, Bender, & Lachmann, 2002).

Experimentalists who hold such a view could think that even if philosophers would share with others the natural and epistemically unproblematic capacity to attribute beliefs, perhaps the attribution of knowledge is a further operation drawing on some pool of mental resources of dubious value. Here it is useful to observe that psychologists working on mental state ascription explicitly classify both knowledge and belief as mental states (e.g. Apperly, 2011; O'Neill, Astington, & Flavell, 1992; Shatz, Wellman, & Silber, 1983). The attribution of knowledge is not seen as an additional step to be taken after belief has been ascribed, and with the help of auxiliary faculties; if anything it is belief that is regarded as the more difficult and problematic state to attribute (on this point, see especially Apperly, 2011). Some philosophers share the view that knowledge is a mental state in its own right, and not a composite of belief and other factors (e.g. McDowell, 1995; Williamson, 2000). On Williamson's view, for example, the concept of belief is derivative from the concept of knowledge: rather than our having to add extra ingredients to an ascription of belief to get an ascription of knowledge, we in some sense subtract from knowledge to ascribe belief—indeed, 'mere belief is a kind of botched knowing' (Williamson, 2000, 47). The psychological literature on mindreading aligns more closely with this view than with the more widespread philosophical opinion according to which knowledge is not considered a mental state. It might be objected that Williamson has what is from the mainstream philosophical perspective a non-standard understanding of 'mental state', but even if that point were conceded, the larger issue would be untouched: it is uncontroversial in psychology to see intuitive attributions of knowledge as falling under the scope of our mindreading capacity, so anyone who generally trusts that latter capacity has reason to trust intuitive knowledge ascriptions.

If intuitive knowledge ascriptions for epistemological cases do arise from our ordinary mindreading capacity, we need to explain reported patterns of variation in responses to Gettier cases and the like. A review of the evidence is in order.

In their influential 2001 paper on the case method, Weinberg, Nichols and Stich discussed the results of polling undergraduates on the following scenario:

> Bob has a friend, Jill, who has driven a Buick for many years. Bob therefore thinks that Jill drives an American car. He is not aware, however, that her Buick has recently been stolen, and he is also not aware that Jill has replaced it with a Pontiac, which is a different kind of American car. Does Bob really know that Jill drives an American car, or does he only believe it?
>
> REALLY KNOWS   ONLY BELIEVES

18

For this case (the only Gettier case tested by Weinberg, Nichols and Stich), 74% of participants who self-identified as "Western" had the standard knowledge-denying response; however, among the East Asian subjects only 43% had this response, and among the South Asian subjects only 39% did. On two further epistemological cases (both 'skeptical pressure' cases in which possibilities of error were mentioned but stipulated to be non-actual) the responses of Western and East Asian subjects were not significantly different, but South Asian subjects were still out of line, responding at levels of 70 and 50% where Western subjects registered 89 and 69% (Weinberg et al., 2001). These are statistically significant differences in performance. But several factors raise questions about whether these results are best interpreted as pointing to underlying differences in intuitive competence. One might notice that if we take the standard "Western" responses as normative, the non-normative responses lie closer to the 50-50 split that one sees when subjects are not interested in a problem and are just answering randomly. In fact, the cases that show no difference between Western and East Asian subjects both produce similar drops towards randomness in the South Asian participants. It is not clear why. One possible explanation here would involve differences in motivation: it is possible that the 24 South Asian participants in the study were on average simply less engaged with these particular problems. If for example South Asian students on the campus where this research was conducted were significantly less likely to be humanities majors,[11] and if participants were thinking of the series of puzzles as an exercise in applying philosophical methods, then the prospect of pondering these cases may have been less attractive to them. One might also wonder whether the content of the Bob and Jill scenario resonated more with members of some ethnic groups than others, or whether some groups found it easier to read.[12]

For a more reliable measure of variation in responses to Gettier cases, it is useful to examine responses to multiple cases. In a new study involving 222 undergraduate participants from a variety of ethnic backgrounds (71 participants self-indentified as White; 58 as South Asian, 28 as East Asian[13], 16 Latin American, 15 Black, 6 West Asian, 4 Arab and 24 other), eight different Gettier

---

[11] I do not have data for Rutgers University in particular, but in the United States, there are statistically significant differences in the college majors of different ethnic groups (e.g. compared to White students, Asian students were at the time of Weinberg, Nichols and Stich's study more than twice as likely to major in biology or engineering). Source: National Center for Education Statistics, *Status and Trends in the Education of Racial and Ethnic Minorities*, table 25.2, Percentage of degrees conferred by degree-granting institutions in most popular fields of study, by race/ethnicity and level of study: 2003-04.

[12] For example, the story will be more easily understood by readers for whom it is an automatic inference that a Buick is an American car. Thanks to Meredyth Daneman for this observation.

[13] Participants in our East Asian group actually identified themselves as Chinese, Japanese, Korean, Filipino or Southeast Asian (Cambodian, Vietnamese, etc.) following the standard Canadian census categories.

cases were tested, together with eight ordinary knowledge cases, eight justified false belief (JFB) cases, eight skeptical pressure cases and eight filler questions involving various types of justified and unjustified belief (Nagel, San Juan, & Mar, in prep). The 32 experimental items were 'matched': for each Gettier case, there was a corresponding knowledge case, skeptical pressure case and JFB case with an equal word count. These items were distributed in a between-subjects design, so that each participant saw just one member of each set of matched items (for example, one of A-D below); each participant judged two experimental items of each kind.

(A). **[Ordinary knowledge:]** Wanda is out for a weekend afternoon walk. She lives in a large new condominium tower downtown, and her suite is fairly small and does not have any windows that open, so she really likes to get out for some fresh air. Passing near the train station, Wanda wonders what time it is. She glances up at the clock on the train station wall and sees that it says 4:15 pm. It is in fact 4:15 pm at that moment.

(B). **[Skeptical pressure variant:]** Wanda is out for a weekend afternoon walk near the train station and wonders what time it is. She glances up at the clock on the train station wall and sees that it says 4:15 pm. It is in fact 4:15 pm at that moment. The station clock is in fact working, but it has no second hand, and Wanda only looks at it for a moment, so she would not be able to tell if the clock were stopped.

(C). **[Gettier case variant:]** Wanda is out for a weekend afternoon walk. As she passes near the train station, she wonders what time it is. She glances up at the clock on the train station wall and sees that it says 4:15 pm. What she doesn't realize is that this clock is broken and has been showing 4:15 pm for the last two days. But by sheer coincidence, it is in fact 4:15 pm just at the moment when she glances at the clock.

(D). **[JFB variant:]** Wanda is out for a weekend afternoon walk. She lives in a small condo downtown, but enjoys the outdoors. As she passes near the train station, she wonders what time it is. She glances up at the clock on the train station wall and sees that it says 4:15 pm. It is in fact 4:53 pm at that moment. What she doesn't realize is that this clock is broken and has been showing 4:15 pm for the last two days.

As each scenario was presented on screen, participants were asked to verify the key proposition in the story (e.g. "According to the story, what time is it when Wanda looks at the clock?"). Participants who answered this question correctly were then asked about Wanda's state of mind (the scenario remained on screen throughout). Unsurprisingly, participants were most likely to ascribe knowledge in the type-A baseline stories (overall, 72.0% agreed that the subject knew the key proposition in such cases), and least likely to ascribe knowledge in the type-D false belief stories (15.8%). Gettier cases and skeptical pressure cases were seen by most participants as involving a failure to know, attracting knowledge ascription rates of 32.9 and 39.8% respectively. Not everyone produced the standard response to every Gettier case, but differences between standard and non-standard responses were not a function of ethnicity or gender. We did not find

statistically significant correlations between ethnicity or gender and knowledge ascription for any of the four key types of scenario.[14]

If undergraduates with little or no philosophical training[15] generally classify epistemology cases the way epistemologists would, and if there is no appreciable gender or ethnic variation in case responses, then there is no good reason to doubt the default hypothesis that we are relying on a common mindreading capacity in responding to these cases. If I am right that our knowledge-denying intuitive responses to skeptical pressure cases are natural illusions of the mindreading capacity (Nagel, 2010), then the evidence that we are responding to epistemic cases on a similar basis is stronger still: the untrained share not only our ordinary responses but also share at least one important illusion.

Critics of the case method may however worry that the signal from the lay responses is not as strong as it should be. If epistemologists have clear intuitions that Gettier cases involve a failure to know, then one might have expected a higher level of consensuality in amateur judgments of those cases. Our amateurs generally agreed with the standard professional line on the eight core Gettier cases we tested, but at a perhaps disheartening average rate of 67.1% (to say nothing of the 15.8% of participants willing to ascribe knowledge in JFB cases).

There are some mundane reasons why one might find a worse signal-to-noise ratio in lay judgments of epistemological scenarios. While intuitive judgment itself is effortless, effort is required to read the stories closely enough to register the relevant details about the inputs to the subject's judgment (cf. Williamson, 2011). Participants who are not interested in a particular story may be more inclined to respond to it randomly.[16] Philosophers and others may have the same

---

[14] For Gettier cases in particular, the correlation between ethnicity and knowledge ascription was 0.018, p=.791 (two-tailed), n=222; the correlation between gender and knowledge ascription was 0.019, p=.774 (two-tailed), n=222. The p-value or significance level indicates the probability of the observed patterns on the assumption that the null hypothesis is true (i.e. that ethnicity/gender does not influence knowledge ascription). Correlations are not regarded as significant in psychology if the significance level is greater than .05. Wright also found no gender differences; she does not report on ethnicity (Wright, 2010).

[15] 58 out of our 222 participants (26%) reported having taken at least one philosophy course; of these 39 reported having taken only one such course. Knowledge ascription rates were not correlated with the number of philosophy courses taken, a result consistent with Jennifer Cole Wright's studies of responses to cases like Lehrer's Truetemp and the Fake Barn Gettier case (Wright, 2010).

[16] One of the filler questions completed by all 222 participants was a verbatim copy of the Jill/Bob Pontiac and Buick case from (Weinberg et al., 2001). We did not reproduce the original finding that South Asians tended to give the non-standard response: only 40% of our South Asian participants ascribed knowledge in this case (vs. 61% of South Asians in Weinberg, Nichols and Stich's original study). We did however see a weaker response to this particular case among South Asians than among self-identified Whites: only 14% of our White participants ascribed knowledge in this case (vs. 26% of 'Western' participants in the original

basic intuitive capacity to register the presence or absence of knowledge, but philosophers may be more motivated to read epistemic vignettes with an eye to exercising this capacity. If our capacity to recognize knowledge is intuitive, we do not necessarily have to devote personal-level attention to the factors justifying our judgments: for example, even if reliability is a necessary condition of knowing, we would not need to make explicit personal-level judgments of reliability in order to see someone as knowing. However, some interest in the relevant parts of the story may be needed in order to activate the relevant spontaneous processing: participants who skim or read inattentively may retain only a rough sense of the gist of the story when they reach the question concerning mental state. Gettier cases can be taxing to follow: amateur participants who do not have the motivation of testing an epistemological theory may be less inclined to read closely.[17]

Motivation may also play some role in the other type of variation that has been considered problematic for the case method. Experimentalists have shown that responses to philosophical scenarios exhibit order or contrast effects: for example, naïve subjects are significantly more likely to attribute knowledge to the subject of a version of Keith Lehrer's Truetemp case when it is presented following a very clear case of ignorance than when it follows a very clear case of knowledge (Swain et al., 2008; Wright, 2010). Intended as a counterexample to reliabilism, this case involves a person whose brain has been surreptitiously changed to produce accurate beliefs about the temperature without his being aware of the change; many epistemologists have the intuition that these reliably true beliefs do not constitute knowledge. Observing that the case may or may not appear to be a case of knowledge depending on its context, Swain and collaborators

---

study). Because we did not see any significant correlation between ethnicity and knowledge ascription across the eight other Gettier cases tested, it seems more reasonable to attribute the difference in response rates here to something like differing levels of interest or engagement with this particular story as opposed to differences in knowledge ascription for Gettier cases per se. I do not know enough about the testing conditions of the original study to speculate about why we might have seen a clearer signal from both groups in our study.

[17] Contingent differences in motivation, as opposed to underlying cognitive differences, have been posited as explaining many of the findings on cognitive 'cross-cultural differences' that served as the theoretical basis of Weinberg, Nichols and Stich's claim of diversity in epistemic intuitions. Richard Nisbett and collaborators maintain that there are fundamental qualitative differences in the reasoning of Eastern and Western peoples, ascribing an "intuitive", "holistic" and "experience-based" way of thinking to those on the Eastern side, and an "analytic", "rule-based" and "decontextualized" mode of thought to the Westerners (Nisbett, Peng, Choi, & Norenzayan, 2001). Hugo Mercier observes that these generalizations do not square well with the historical record of Chinese work on logic. He argues that Nisbett's Eastern subjects have given less analytic—and more shallow—responses because these subject pools happened to be less interested in Nisbett's tasks (Mercier, forthcoming). There is evidence that differences are not found when care is taken to ensure that motivation is uniform: for example, even just imagining that one has a stake in some issue under contention can suffice to erase cultural differences in reasoning between Eastern (Japanese) and Western (French) subjects (Van der Henst, Mercier, Yama, Kawasaki, & Adachi, 2006).

contend that variation of this kind undermines the case method: they are concerned that sensitivity to what one has recently been thinking renders intuitions disturbingly open to manipulation and only questionably connected to the subject matter of interest.

The possibility of contextual influences on some capacity does not on its own show that this capacity cannot be a source of evidence. For example, one finds contrast effects in visual judgments of color:
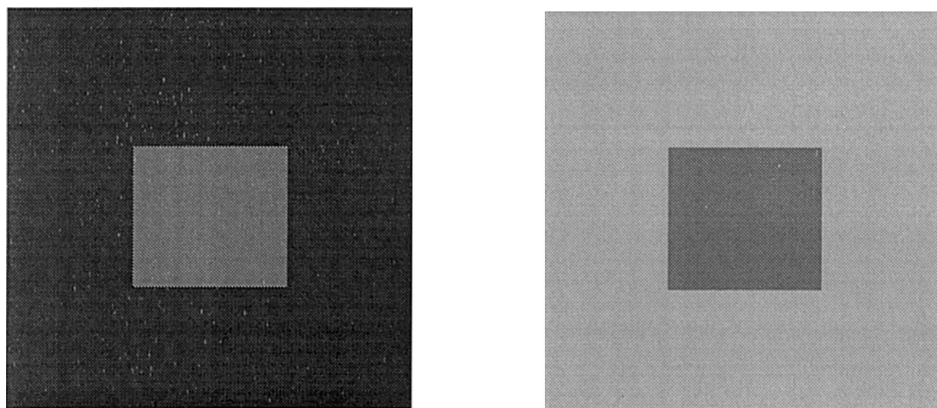
Swain et al. explicitly consider the objection that the types of variability they have identified are also found in perception. They contend there is a relevant disanalogy: "We are aware of the great majority of the circumstances under which perceptual judgments are likely to be unreliable. For instance, we know that visual perception requires a certain amount of illumination, and visual perception itself provides us with knowledge of whether enough illumination is present" (Swain et al., 2008, 148). Swain et al. observe that we naturally have lower confidence in perceptual judgments in conditions of low illumination; the defender of the case method might observe that we naturally have lower confidence in borderline intuitive judgments as well, including the version of the Truetemp case that Swain et al. are testing (Wright, 2010). Swain et al. also suggest that correcting for unreliability will be easier in the perceptual case: while we can counter dim illumination in obvious ways, "we don't know what is the parallel for intuition of making sure that the light is on; that is, we do not know which are the circumstances that render intuition reliable or unreliable" (ibid.).

It is doubtless easy enough to keep the light on, but it is not clear that low illumination is a good parallel for the order effects identified for the Truetemp case. A closer analogy for order effects from the domain of visual perception would be order effects in color judgments:

identification of colors is sensitive to what other colors we have been seeing, so that we will describe the same color patch as green if we approach it from one end of the spectrum and as blue if we approach it from the other (Kalmus, 1979).[18] These kinds of contextual effects do not announce themselves as obviously as low illumination does, and it is not a trivial matter to figure out how we should assess them, whether such effects should be seen as occasioning errors in our application of color concepts, or as showing that the boundaries of our concepts are more variable than we might have supposed (Raffman, 2005).   Naïve perception on its own does not tell us what to do here.

Meanwhile, correcting for order effects in the case method might not be insuperably difficult. Simon Cullen has argued that the order effects found by Swain et al. arise from adherence to conversational norms in the survey context: participants noticed the obvious contrast between the Truetemp and extreme cases of knowledge and ignorance, and understood the question about Truetemp's state of mind as an invitation to compare him with the salient contrasting case. On this view, participants given a different order of cases would have a different conception of their task. Cullen tested the pairs of cases that generated the strongest order effects for Swain et al., but this time with instructions urging participants to "consider each independently." Clarifying the pragmatics in this manner was enough to cancel the order effects: with the new instructions no significant variation arose from presenting the Truetemp case after a clear case of knowledge as opposed to a clear case of ignorance (Cullen, 2010). If epistemologists are generally making an effort to consider their cases independently, the Swain results should not be cause for alarm.

The Truetemp studies may raise worries of another kind, however. As Swain et al. observe, most epistemologists tend to have the intuition that Truetemp does not know the temperature: even reliabilists like Alvin Goldman see the case as prima facie intuitive evidence against reliabilism (Goldman, 1994). Both Cullen and Swain found considerable ambivalence about this case among amateurs, however, with mean responses of about 2.5 to 3.3 on a 1-5 scale where 3 was neutral and 4 or 5 would indicate a denial or strong denial of knowledge; Wright's results were more negative but also failed to show strong consensus on the case. One might worry that differences between amateurs and professionals here point to the involvement of some capacity other than the universal mindreading capacity in deciding whether Truetemp has knowledge.

It is possible that generic motivation problems played a role here: perhaps the case is confusing enough that amateur participants give up on following what is happening and respond near the midpoint simply to express uncertainty about the task (cf. De Bruin, Fischhoff, Millstein, &

---

[18] Thanks to Diana Raffman for the example and discussion of the point.

Halpern-Felsher, 2000). Another possibility is that the case is somewhat under-described, and amateur participants are not all fleshing out the details the same way (as each other, or as professionals would). Here is the version of the case used by Swain, Cullen and Wright:

> One day Charles was knocked out by a falling rock; as a result his brain was ''rewired'' so that he is always right whenever he estimates the temperature where he is. Charles is unaware that his brain has been altered in this way. A few weeks later, this brain rewiring leads him to believe that it is 71 degrees in his room. Apart from his estimation, he has no other reasons to think that it is 71 degrees. In fact, it is 71 degrees.

Just given this text, there are various ways to construe Charles's predicament. One might wonder, in particular, whether he is in any way self-conscious about the peculiarity of his obviously peculiar new tendency to form precise beliefs about the temperature. Keith Lehrer's original presentation of the example—involving a 'tempucomp' implanted by a brain surgeon—provided more detail on this score. In Lehrer's version, Truetemp is "slightly puzzled about why he thinks so obsessively about the temperature". Lehrer goes on to stipulate that Truetemp "never checks a thermometer to determine whether these thoughts about the temperature are correct. He accepts them unreflectively, another effect of the tempucomp" (Lehrer, 1990, 163). One of the interesting features of Lehrer's version is that the tempucomp is described as having two distinct effects: first, the production of accurate temperature thoughts in Truetemp's mind, and second, the production of Truetemp's personal-level but somehow unreflective acceptance of these thoughts.[19] When we are reconstructing Truetemp's mental state, we go through a two-step process, where the second step (in which the temperature thoughts are accepted without reason) is problematic, perhaps even from a reliabilist perspective. If we naturally see the tempucomp as mechanically disabling Truetemp's ordinary capacity to weigh evidence as he decides what to accept, and if we see that ordinary capacity as generally conducive to the production of true belief, then the tempucomp intervention could well come across as the type of thing that would generally compromise reliability.[20]

---

[19] On the thought that acceptance is a personal-level phenomenon, see (Frankish, 2009). A similar two-stage structure is used in another much-discussed internalist case developed by Laurence BonJour: his unwitting clairvoyant not only forms various accurate beliefs through her psychic faculty, but also, in a distinct step, "accepts the beliefs in question" (BonJour & Sosa, 2003, 28).

[20] On this reading, the case is not in the end an effective counterexample to reliabilism. For present purposes it is enough that the case is ambiguous, but more negative in its original Truetemp form; a clear decision about what type of process is intuitively seen as instantiated would require grappling with the notoriously difficult Generality Problem (Conee & Feldman, 1998), not a task I will undertake here.

The simpler Charles example does not dwell on the separation between the generation and the acceptance of propositions about temperature: rather, Charles's brain re-wiring leads directly to a belief. Where the original Truetemp seemed to be making dubious reflective judgments, perhaps Charles comes across as making better intuitive judgments. But the new case does not strictly preclude a more negative reading: when we read that Charles has 'apart from his estimation... no other reasons to believe that it is 71 degrees', it is open to us to represent him as actively thinking to himself that he has no reasons for his belief, and if we see him as engaged in that type of reflection on the way to maintaining his belief, then his judgment once again appears unreliable. This more reflective construal of the Charles case may be more available to professionals familiar with the similarly negative original case.

Where there are multiple ways of understanding an under-described story, better consensus among professionals may be a result of our naturally favoring the construal that makes more dialectical sense in its original argumentative context.[21] Someone whose way of fleshing out a case is non-standard could have strong confidence in what seems to be a minority intuition; divergent intuitions on a given case in epistemology could in some circumstances arise from different ways of construing various subtle features of the subject's way of thinking, rather than differences in the assessment of knowledge per se.[22] It is doubtless a risk of the case method that we can fill in under-described cases with philosophically important content.[23] The problem is not an intractable one, however: disagreement about cases can help us to identify this content. The problem is also not unique to the case method: perception can also present us with stimuli that are ambiguous between two or more construals, stimuli such as the Necker cube.

So far, the aim has been to show that epistemic intuitions do not show peculiar or particularly problematic forms of variation. The kinds of instability found in epistemic intuitions are also found in perceptual judgments. Amateur judgments of ordinary knowledge, Gettier,

---

[21] This possibility is particularly relevant to our understanding of the paired cases motivating contextualism and interest-relative invariantism, where it is clear that the point is that there is *some* way of reading the pair so that they will deliver contrary verdicts. Cooperative readers can then work to construe the cases in such a way. To motivate contextualism or IRI, these ways of reading the cases do not necessarily have to be the ways that one would most naturally read these cases if encountering them in isolation.

[22] It is possible that the thinking of the subject in the Fake Barn case is also open to being understood in various significantly different ways as well, and that this under-description also explains lukewarm amateur responses and some division in professional responses to the case.

[23] The importance of tacit commitments in our understanding of cases has been explored most systematically by Tamar Gendler: see the essays collected in (Gendler, 2011).

skeptical pressure and false belief cases align roughly with the judgments of professionals, and do not seem to vary with ethnicity or gender. There are no obvious barriers to seeing epistemic case intuitions as the products of our ordinary mindreading capacities.

A last problem concerns our handling of illusions. Mindreading, like perception and like other intuitive capacities, is susceptible to certain natural illusions, most notably epistemic egocentrism, or the tendency to misrepresent the cognition of those who occupy a more naïve position than the observer (e.g. Birch & Bloom, 2004; 2007; Nickerson, 1999). The legitimacy of the case method does not require that philosophers be immune to these illusions, or able to distinguish illusory from non-illusory cases immediately as they are experienced. Even advanced reflection and training does not insulate one from illusion: for example, physics graduate students and postdoctoral researchers still experience the characteristic cognitive-perceptual illusions of naïve 'impetus theory' physics, illusions that are explicitly at odds with their theoretical knowledge and easily discounted on reflection (Kozhevnikov & Hegarty, 2001).

Fortunately, in epistemology as in physics, intuition is not the only tool at our disposal: considerations of theoretical unification can also supply some guidance.[24] We suspect sensory illusion where the deliverances of the senses appear to conflict with one another, as in the Müller-Lyer illusion; we may have similar suspicions where there is apparent conflict among our epistemic intuitions, for example, conflict of the sort found in the cases motivating contextualism. It is not transparent that these apparently conflicting intuitions are illusions; theorists of various inclinations have developed innovative and sometimes strange theories of knowledge and knowledge ascription on which the apparent conflict is no more than apparent. To judge whether these theories are true or false, we can draw on a great array of considerations from logic, linguistics, psychology and philosophy; we can also devise new cases to offer positive support to our theories or to serve as counterexamples. In epistemology, as in empirical science, it is not always a trivial matter to determine whether we are subject to an illusion, or whether the phenomenon we are investigating is stranger than we had thought. But the fact that a form of inquiry is difficult does not entail that there is anything fundamentally wrong with its methods.[25]

---

[24] For a more detailed discussed of this point, see (Ichikawa, forthcoming).
[25] For comments on an earlier draft I am grateful to Jane Friedman, Diana Raffman, Sergio Tenenbaum, Jonathan Weinberg and Jonathan Weisberg.

References

Alexander, J., & Weinberg, J. M. (2007). Analytic epistemology and experimental philosophy. *Philosophy Compass, 2*(1), 56-80.

Alter, A., & Oppenheimer, D. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review, 13*(3), 219.

Alter, A., Oppenheimer, D., Epley, N., & Eyre, R. (2007). Overcoming intuition: Metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology: General, 136*(4), 569-576.

Apperly, I. (2011). *Mindreaders: The Cognitive Basis of "Theory of Mind"*. Hove and New York: Psychology Press.

Apperly, I., Back, E., Samson, D., & France, L. (2008). The cost of thinking about false beliefs: Evidence from adults' performance on a non-inferential theory of mind task. *Cognition, 106*(3), 1093-1108.

Apperly, I., Riggs, K., Simpson, A., Chiavarino, C., & Samson, D. (2006). Is belief reasoning automatic? *Psychological Science, 17*(10), 841.

Baranski, J. V., & Petrusic, W. M. (1994). The calibration and resolution of confidence in perceptual judgments. *Attention, Perception, & Psychophysics, 55*(4), 412-428.

Baranski, J. V., & Petrusic, W. M. (1995). On the calibration of knowledge and perception. *Canadian Journal of Experimental Psychology, 49*(3), 397-407.

Birch, S., & Bloom, P. (2004). Understanding children's and adults' limitations in mental state reasoning. *Trends in cognitive sciences, 8*(6), 255-260.

Birch, S., & Bloom, P. (2007). The curse of knowledge in reasoning about false beliefs. *Psychological Science, 18*(5), 382-386.

BonJour, L., & Sosa, E. (2003). *Epistemic Justification: Internalism vs. Externalism, Foundations vs. Virtues*. Malden, MA: Blackwell.

Brewer, W. F., & Sampaio, C. (2006). Processes leading to confidence and accuracy in sentence recognition: A metamemory approach. *Memory, 14*(5), 540-552.

Buckwalter, W., & Stich, S. (2011). Gender and the Philosophy Club. *The Philosophers' Magazine*(52), 60-65.

Cesarini, D., Sandewall, & Johannesson, M. (2006). Confidence interval estimation tasks and the economics of overconfidence. *Journal of Economic Behavior & Organization, 61*(3), 453-470.

Charman, T., Ruffman, T., & Clements, W. (2002). Is there a gender difference in false belief development? *Social Development, 11*(1), 1-10.

Choi, I., & Nisbett, R. E. (2000). Cultural psychology of surprise: Holistic theories and recognition of contradiction. *Journal of personality and social psychology, 79*(6), 890-905.

Chudnoff, E. (2010). The nature of intuitive justification. *Philosophical Studies, 153*(2), 313-333.

Cohen, A., & German, T. (2009). Encoding of others' beliefs without overt instruction. *Cognition, 111*(3), 356-363.

Conee, E., & Feldman, R. (1998). The generality problem for reliabilism. *Philosophical Studies, 89*(1), 1-29.

Cullen, S. (2010). Survey-driven romanticism. *Review of Philosophy and Psychology, 1*, 275-296.

De Bruin, W. B., Fischhoff, B., Millstein, S. G., & Halpern-Felsher, B. L. (2000). Verbal and Numerical Expressions of Probability. *Organizational Behavior and Human Decision Processes, 81*(1), 115-131.

DeRose, K. (2009). *The Case for Contextualism: Knowledge, Skepticism, and Context, Volume 1*. New York: Oxford University Press.

Ekman, P. (1989). The argument and evidence about universals in facial expressions of emotion. In H. Wagner & A. Manstead (Eds.), *Handbook of social psychophysiology* (pp. 143-164). Oxford: John Wiley and Sons.

Ekman, P., & Friesen, W. V. (1975). *Unmasking the face: A guide to recognizing emotions from facial clues*: Prentice-Hall Englewood Cliffs, NJ.

Evans, J. (2007). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology, 59*, 255-278.

Frankish, K. (2009). Systems and levels: dual-system theories and the personalsubpersonal distinction. *In Two Minds: Dual Processes and Beyond*, 89-107.

Gendler, T. S. (2011). *Intuition, Imagination, and Philosophical Methodology*. New York: Oxford Univ Pr.

Gettier, E. L. (1963). Is Justified True Belief Knowledge? *Analysis, 23*, 121-123.

Gigerenzer, G., Hoffrage, U., & Kleinbolting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological review, 98*(4), 506-528.

Goldman, A. (1967). A Causal Theory of Knowing. *The Journal of Philosophy, 64*(12), 357-372.

Goldman, A. (1976). Discrimination and Perceptual Knowledge. *The Journal of Philosophy, 73*(20), 771-791.

Goldman, A. (2006). *Simulating minds: The philosophy, psychology, and neuroscience of mindreading.* New York: Oxford University Press.

Goldman, A. I. (1994). Naturalistic epistemology and reliabilism. *Midwest Studies in Philosophy, 19*(1), 301-320.

Heine, S. J., & Lehman, D. R. (1996). Hindsight bias: A cross-cultural analysis. *Japanese Journal of Experimental Social Psychology, 35*, 317-323.

Ichikawa, J. (forthcoming). Experimentalist pressure against traditional philosophical methodology. *Philosophical Psychology.*

Kalmus, H. (1979). Dependence of colour naming and monochromator setting on the direction of preceding changes in wavelength. *The British journal of physiological optics, 33*(2), 1.

Kelley, C. (1993). Remembering mistaken for knowing: Ease of retrieval as a basis for confidence in answers to general knowledge questions, *Journal of Memory and Language* (Vol. 32, pp. 1).

Koriat, A. (1975). Phonetic symbolism and feeling of knowing. *Memory & Cognition, 3*(5), 545-548.

Koriat, A. (1976). Another look at the relationship between phonetic symbolism and the feeling of knowing. *Memory & Cognition, 4*(3), 244-248.

Koriat, A. (1995). Dissociating knowing and the feeling of knowing: Further evidence for the accessibility model. *Journal of Experimental Psychology: General, 124*(3), 311-333.

Koriat, A. (2008). Subjective confidence in one's answers: The consensuality principle. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(4), 945-959.

Koriat, A. (2011). Subjective Confidence in Perceptual Judgments: A Test of the Self-Consistency Model. *Journal of Experimental Psychology: General, 140*(1), 117-139.

Koriat, A., & Adiv, S. (in press). The construction of attitudinal judgments: Evidence from attitude certainty and response latency. *Social Cognition.*

Kozhevnikov, M., & Hegarty, M. (2001). Impetus beliefs as default heuristics: Dissociation between explicit and implicit knowledge about motion. *Psychonomic Bulletin & Review, 8*(3), 439.

Lehrer, K. (1990). *Theory of Knowledge.* Boulder: Westview Press.

Lindsey, D., Raffman, D., & Brown, A. (MS). Psychological hysteresis and the nontransitivity of insignificant differences.

Mar, R. A., Kelley, W. M., Heatherton, T. F., & Macrae, C. N. (2007). Detecting agency from the biological motion of veridical vs animated agents. *Social cognitive and affective neuroscience, 2*(3), 199.

McDowell, J. (1995). Knowledge and the Internal. *Philosophy and Phenomenological Research, 55*(4), 877-893.

Mercier, H. (forthcoming). On the universality of argumentative reasoning. *Journal of Cognition and Culture.*

Mercier, H., & Sperber, D. (2009). Intuitive and reflective inferences. *J. St. BT Evans & K. Frankish (Eds.), In two minds: Dual processes and beyond. Oxford, UK: Oxford University Press.*

Nagel, J. (2007). Epistemic intuitions. *Philosophy Compass, 2*(6), 792-819.

Nagel, J. (2010). Knowledge ascriptions and the psychological consequences of thinking about error. *Philosophical Quarterly, 60*(239), 286-306.

Nagel, J. (2011). The psychological basis of the Harman-Vogel paradox. *Philosophers' Imprint, 11*(5), 1-28.

Nagel, J., San Juan, V., & Mar, R. (in prep). Gettier Case Recognition.

Newell, B., & Lee, M. (2010). The Right Tool for the Job? Evidence Accumulation in Decision Making. *Psychological review.*

Nickerson, R. S. (1999). How we know--and sometimes misjudge--what others know: Imputing one's own knowledge to others. *Psychological Bulletin, 125*(6), 737-759.

Nisbett, R. E., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: Holistic versus analytic cognition. *Psychological Review, 108*(2), 291-310.

O'Neill, D., Astington, J., & Flavell, J. (1992). Young children's understanding of the role that sensory experiences play in knowledge acquisition. *Child Development, 63*(2), 474-490.

Pohl, R. F., Bender, M., & Lachmann, G. (2002). Hindsight bias around the world. *Experimental Psychology, 49*(4), 270-282.

Raffman, D. (2005). How to understand contextualism about vagueness: reply to Stanley. *Analysis, 65*(287), 244-248.

Reber, R., & Schwarz, N. (1999). Effects of perceptual fluency on judgments of truth, *Consciousness and cognition* (Vol. 8, pp. 338-342). Orlando, Fla.: Academic Press.

Redcay, E., Dodell-Feder, D., Pearrow, M. J., Mavros, P. L., Kleiner, M., Gabrieli, J. D. E., & Saxe, R. (2010). Live face-to-face interaction during fMRI: A new tool for social cognitive neuroscience. *NeuroImage, 50*(4), 1639-1647.

Samson, D., Apperly, I., Braithwaite, J., Andrews, B., & Scott, S. (2010). Seeing it their way. *Journal of Experimental Psychology-Human Perception and Performance, 36*(5), 1255-1266.

Saxe, R. (2006). Why and how to study Theory of Mind with fMRI. *Brain Research, 1079*(1), 57-65.

Saxe, R., & Wexler, A. (2005). Making sense of another mind: The role of the right temporo-parietal junction. *Neuropsychologia, 43*(10), 1391-1399.

Shatz, M., Wellman, H. M., & Silber, S. (1983). The acquisition of mental verbs: A systematic investigation of the first reference to mental state. *Cognition, 14*(3), 301-321.

Simmons, J. P., & Nelson, L. D. (2006). Intuitive confidence: Choosing between intuitive and nonintuitive alternatives. *Journal of Experimental Psychology-General, 135*(3), 409-427.

Slobin, D. I. (1968). Antonymic phonetic symbolism in three natural languages. *Journal of personality and social psychology, 10*(3), 301-305.

Sloman, S. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119*(1), 3-22.

Sosa, E. (2007). Experimental philosophy and philosophical intuition. *Philosophical Studies, 132*(1), 99-107.

Stanley, J. (2005). *Knowledge and Practical Interests*. New York: Oxford University Press.

Stanovich, K., & West, R. (2000). Individual differences in reasoning: implications for the rationality debate? *The Behavioral and brain sciences, 23*(5), 645.

Stanovich, K. E. (2005). *The robot's rebellion: Finding meaning in the age of Darwin*: University of Chicago Press.

Swain, S., Alexander, J., & Weinberg, J. M. (2008). The instability of philosophical intuitions: Running hot and cold on Truetemp. *Philosophy and Phenomenological Research, 76*(1), 138-155.

Van der Henst, J. B., Mercier, H., Yama, H., Kawasaki, Y., & Adachi, K. (2006). Dealing with contradiction in a communicative context: A cross-cultural study. *Intercultural Pragmatics, 3*(4), 487-502.

von Helmholtz, H. (1893). *Popular lectures on scientific subjects*. London: Longmans, Green and co.

Weatherson. (2003). What good are counterexamples? *Philosophical studies, 115*, 1-31.

Weinberg, J. M. (2007). How to challenge intuitions empirically without risking skepticism. *Midwest Studies in Philosophy, 31*, 318-343.

Weinberg, J. S., Nichols, S., & Stich, S. (2001). Normativity and Epistemic Intuitions. *Philosophical Topics, 29*(1), 429-460.

Wellman, H., Cross, D., & Watson, J. (2001). Meta analysis of theory of mind development: The truth about false belief. *Child Development, 72*(3), 655-684.

Williamson, T. (2000). *Knowledge and its Limits*. New York: Oxford University Press.

Williamson, T. (2005). Contextualism, Subject-Sensitive Invariantism and Knowledge of Knowledge. *The Philosophical Quarterly, 55*, 213-235.

Williamson, T. (2007). *The Philosophy of Philosophy*. New York: Wiley-Blackwell.

Williamson, T. (2011). Philosophical Expertise and the Burden of Proof. *Metaphilosophy, 42*(3), 215-229.

Wright, J. C. (2010). On intuitional stability: The clear, the strong, and the paradigmatic. *Cognition, 115*(3), 491-503.

Zagzebski, L. (1994). The inescapability of Gettier problems. *The Philosophical Quarterly, 44*(174), 65-73.