PHILOSOPHICAL NATURALISM AND INTUITIONAL METHODOLOGY

Alvin I. Goldman

Rutgers University

Romanell Lecture on Philosophical Naturalism

American Philosophical Association, 2010

Forthcoming in: Proceedings and Addresses of the American Philosophical Association

1. Intuitions in Philosophy: What's the Controversy?

A debate is raging over philosophical methodology. It is a debate between philosophical traditionalists and science-oriented philosophical naturalists concerning the legitimacy of the widespread use of intuitions in philosophy. Not everyone finds the term 'intuition' the best label for what philosophers rely upon in the relevant sector of their practice. Instead of "intuitions" some prefer to talk of intuitive judgments, thought experiments, or what have you. Nonetheless, "intuition" is the most commonly used term in the territory, so I shall not abandon it, though other terms will be used as well. Described fairly neutrally, the philosophical activity in question consists of specifying a hypothetical scenario and rendering an intuitive judgment about the correctness or incorrectness of classifying it under a stipulated heading. Does a given predicate 'F' apply to an event, an individual, a pair of objects, etc. in the scenario? Putting the question less linguistically, is a certain property or relation exemplified in the scenario? The central question in the debate is whether, when intuitions or intuitive judgments are formed about such cases, they provide good evidence for some type of philosophical conclusion. For instance, are they good evidence for a conclusion of the form "Case C is/is not an instance of F"? Philosophers' ultimate interest is not in cases per se. Cases are examined as a means to determining the content or composition of the referent of 'F' (where the referent may be a property, a kind, a meaning, or a concept). But this further determination is not in the foreground -- though it definitely cannot be ignored.

That case intuitions sometimes have a powerful impact on philosophical thinking is evident from historically salient cases. Gettier's famous paper of 1963 is a prime example. It was concurring intuitive judgments about his counterexamples to the justified true belief analysis of knowledge that propelled major changes in our theory of knowledge, or the analysis of knowledge. Today the JTB analysis is a dead letter. So the impact of intuitive judgments in philosophy cannot be underestimated. Whether philosophers are right to be guided by such judgments remains to be settled.

Let us not worry excessively about how much of philosophy is done in this mold, or how important a facet of philosophy it constitutes. Clearly, philosophical work is not exhausted by reflection on examples and their classification. Philosophy is also in the business of theoretical construction. Nonetheless, a fair bit of analytic philosophy is preoccupied with generating and examining intuitions about cases. Is this practice legitimate? What is it supposed to tell us, and is it a reliable method for generating evidence about some sought-after philosophical conclusions? If so, what are these conclusions, and does the method produce genuinely good evidence (at least in favorable cases)? How is the practice to be understood, and is it methodologically defensible?

Consulting intuitions about imaginary cases is a paradigm of "armchair" philosophy. Whatever else is meant by an "armchair" method, it is certainly intended to contrast with the method of empirical science, which uses "real" experimentation and

observation rather than reflection on hypothetical examples. Philosophical naturalists tend to distrust armchair methods. They embrace empirical science as the paradigm of sound methodology. Philosophy should not be dragged into epistemological disrepute by utilizing suspect faculties or procedures like intuition. Traditionalists reply that intuition is a perfectly sound faculty; it is just an a priori faculty rather than an empirical one. Rationalists are happy to defend intuition, which they see as closely allied to, if not identical with, the faculty of reason. Hence the sharp split, at least in initial instincts, between traditionalists and philosophical naturalists.

The debate has intensified with the advent of experimental philosophy (X-phi). Proponents of the so-called "negative" branch of X-phi, led by Stephen Stich, claim that their experimental studies raise skeptical doubts about the epistemic soundness or robustness of philosophical intuitions. They challenge the evidential status that the bulk of the profession routinely accords to intuitions (Weinberg, Nichols, and Stich, 2001; Machery, Mallon, Nichols, and Stich, 2004; Alexander and Weinberg, 2007; Swain, Alexander, and Weinberg, 2008; Nichols and Knobe, 2007; Alexander, Mallon, and Weinberg, forthcoming).[1] Although experimental philosophers do not speak for all philosophical naturalists, they are a subclass of philosophical naturalists who have raised objections to the epistemic credentials of intuitions. Negative experimental philosophy will therefore play a prominent, though by no means exclusive, role in the present discussion. I shall not attend closely to the details of their findings to date, nor offer an assessment of the scientific quality of their experimental work (which is sometimes criticized). I am more interested in the ways that experimental work can in principle shed light on the epistemological status of philosophical intuitions.

2. Two Types of Evidential Questions: First-order and Second-order Questions

The debate over intuitional methodology is customarily formulated in terms of "evidence" (or evidential quality), a standard term in the epistemological vocabulary, but one less commonly at center stage than "knowledge," "justification" or "rationality." Moreover, there is much less agreement among epistemologists as to what qualifies as evidence or positive evidential status. I shall therefore advance a proposal that suits present purposes.

Before doing this, however, I want to introduce a distinction that will play a crucial organizing role here. This is the distinction between first-order and second-order questions about evidence, in particular, about the evidential status of intuitions or intuitive judgments. As background, I assume that intuitions are occurrent mental states, either intuitive judgments (a species of doxastic states) or non-doxastic states such as intellectual "seemings" or "attractions," which tend to generate intuitive judgments. In either case, they are supposed to be spontaneous or non-inferential states, not states produced by conscious inference.

The following are some sample first-order questions about the evidential status of intuitions: Are intuitions, or intuitive judgments, evidence at all? What is their evidential

quality or strength? For what propositions or hypotheses are they good evidence? For whom are they good evidence (their subjects only, or other people as well)? What kind of evidence do intuitions have or confer: empirical evidence or a priori evidence?

The following are some sample second-order questions about the evidential status of intuitions: Is there evidence (good evidence?) for the first-order evidential status of intuitions? If we don't already have good enough evidence for the first-order evidential status of intuitions, how should we go about gathering evidence about the first-order status of intuitions? What are the most appropriate or most helpful kinds of second-order evidence: empirical or a priori?

Looking ahead, I shall advance the view (qualified in various ways) that the two orders of evidential status for intuitions are of different types. Specifically, intuitions or intuitive judgments may have first-order evidential status of a substantially a priori kind, while their second-order evidential status – the evidence for their evidential status – is mainly of the empirical kind. This bifurcated treatment is only possible, of course, in light of the distinction between first- and second-order questions about evidence. However, I do not commit myself to a firm endorsement of the bifurcated view. It is a tricky matter, as we shall see in the final section. Whether the first-order evidential status of intuitions is of the a priori kind depends partly on how one defines "a priori," and also on what cognitive science tells us about the cognitive processes underlying classification intuitions.

3. The Nature of Evidential States

What is evidence? That is, what kinds of entities or states of affairs constitute pieces of evidence (for something or other), and in virtue of what relation to an appropriate relatum do they so qualify? Kelly (2008) distinguishes between several senses of "evidence," and I shall focus on one of the senses he pinpoints. This is the reliable indicator sense of evidence, in which X is evidence for Y if and only if X is a (fairly) reliable sign, or indicator, of (the truth or existence of) Y. When using the reliable-indicator sense of "evidence," I suggest that we take states of affairs, or facts, as the entities that constitute evidence. For example, there being 743 rings on the trunk of a given tree is (good) evidence for the proposition that the tree is 743 years old. That a column of mercury in a given thermometer is at the 70-degree mark is (good) evidence that the ambient temperature is 70 degrees Fahrenheit. That a clock reads "4:45" may be (good) evidence that the current time at the clock's location is 4:45. In all of these cases the specified state of affairs constitutes evidence for the specified relatum because the indicated state of affairs is a rather reliable indicator of the truth of the proposition specified.[2]

Can a mental state qualify as good evidence under the proposed sense of evidence? Certainly, at least if we slightly tweak the proposal to read that somebody's being in a certain mental state is the item of evidence. For some types of contentful mental states, it is very plausible that being in one of those states reliably indicates the truth of its content. Seeming to see an apple on a nearby table may be a reliable indicator

of there being an apple on a nearby table. Seeming to remember eating granola for breakfast may be a reliable indicator that one did eat granola for breakfast.

Not all mental states with contents, however, are reliable indicators of the truth of their contents. Wishing that p, imagining that p, fearing that p, and so forth are unlikely to be reliably accompanied by the truth of their contents. However, intuiting that p may be one of the mental states that is a reliable indicator of the truth of its contents, at least for a suitably delimited class of intuitings. If so, such intuitings would qualify as evidential states in the reliable-indicator sense.

Some philosophers argue that this slant on the evidential status of thought experiments is in danger of "psychologizing" the evidence. Williamson (2007) warns against this approach and points to a number of hazards. The evidence extracted from philosophical thought experiments should not be construed as the mental states philosophers undergo, i.e., their intuitings. Rather, the evidence consists in the contents of those intuitings, i.e., the (objective) facts or states of affairs intuited.[3] When a philosopher judges that a character in a Gettier case fails to know proposition p, the item of evidence isn't the philosopher's intuitive judgment; it is the (truth of) the proposition so judged, namely, that the character doesn't know p. Williamson contends that the psychological view (A) misdescribes philosophical practice, (B) rests on a false principle about evidence that he calls "evidence neutrality," and (C) opens up an epistemologically unfortunate gap between thought-experimental evidence and what it is supposed to be evidence for.

There is no room to respond to these worries in detail, but none of them, in my opinion, provides a well-founded reason to reject a psychological approach to intuitional evidence.[4] With respect to the threatened "gap," in particular, the epistemological literature contains a diverse array of theories – of both internalist and externalist varieties – for closing this gap. Moreover, it deserves emphasis that the choice of "intuitings" as the basic kind of evidence underpinning philosophical practice does not exclude "intuiteds" as a derivative species of evidence that can also be invoked. Unquestionably, philosophers appeal to the singular facts they discern through thought experiments as part of their evidence for or against various theories. But such non-psychological evidential facts derive their evidential status (assuming they have such status) from psychological evidential facts, i.e., the occurrence of intuitions or intuitive judgments about the hypothetical scenarios.

4. Indicator Reliability and Process Reliability

Our chosen criterion of evidencehood is the reliable-indicator criterion. But reliable indicatorship is not unrelated to reliable processes. Specifically, processes that lead to a state often contribute to its evidential or non-evidential status. Consider our example of the clock that reads 4:45. Is this reading a reliable indicator of the true time (in the time-zone where the clock is located)? This depends on prior events and processes that led to the clock's current reading. For example, when the clock was last set, was it set correctly relative to the time-zone it was in? If so, has it remained in the

same time-zone until now or has it been moved to a different time-zone? Finally, how well does the clock's time-keeping mechanism work? A defective mechanism, obviously, can make a big difference to the clock's current accuracy. In short, past events and processes that generate the clock's current reading bear on the question of whether the current reading indicates the correct time (in the current time-zone). Hence, in weighing the evidential status of the clock's position vis-à-vis the hypothesis that the current time is 4:45, evidence about those past events and processes is highly germane. If distorting events or processes are known to have transpired, they can constitute (second-order) evidence to the effect that the clock's current reading is poor evidence – if evidence at all -- of the correct time.

For reasons such as this, items that are claimed to be evidence can be examined for bona fide evidential status by investigating their history of generation (among other things). People can evaluate the first-order evidential status of a state or event by considering the processes that led to it. In such a scenario, causal processes provide (one species of) second-order evidence about the first-order evidential status of a given state or event.

Consider forensic evidence presented in a legal context. A standard specimen of forensic evidence is courtroom testimony by a forensic scientist to the effect that a pattern of fingerprints found at a crime scene constitutes a "match" with a defendant's fingerprints. Is this kind of testimony sufficiently reliable and hence worthy of being admitted into court? To ask this question is already to ask a second-order question about the evidence. It is to pose the question of whether there is good evidence that the method of fingerprint matching, as customarily done by forensic scientists, generates testimonial acts that reliably indicate whether fingerprints found at crime scenes are those of the defendant.

Such second-order questions have recently been raised in a very critical way in American courts of law. The so-called "ACE-V" method of fingerprint evidence has long been the gold standard of forensic evidence. But several high-profile cases have occurred in the last number of years in which alleged "matches" generated by the F.B.I. crime laboratory -- long touted as one of the best forensic laboratories -- turned out to be badly mistaken. This led academic specialists to dig deeper into this method. On closer inspection, what the community of fingerprint evidence specialists call "scientific" does not pass muster as very scientific (Mnookin, 2009). (Simply calling a method "scientific" does not make it so.). Moreover, the method's accuracy has never been subjected to proper statistical tests. So we have no scientific second-order evidence that this so-called evidence is good evidence.

The forensic science example illustrates the relevance of generating processes to the potential evidential status of a generated outcome. The example also illustrates the point that being taken to be evidence does not make something genuine evidence, at least not good evidence. If a state of affairs does not reliably indicate the sorts of facts it purports to indicate, it should not be relied upon as evidence. The forensic science example also illustrates the point that empirical tests can be often be designed to probe the reliability of a process or method, in order to decide whether its outputs are reliable indicators of that for which they purport to be

indicators. These tests would be examples of second-order evidence. Finally, just as the legal system should surely demand good second-order evidence about the reliability of courtroom forensic testimony, it is reasonable for philosophers to seek analogous tests to obtain second-order evidence about the evidential quality of philosophical methods.

5. The Negative Program of Experimental Philosophy

Exactly what experimental evidence challenges the reliability of intuitions, specifically, singular classification intuitions? Singular classification intuitions are intuitive judgments that arise when a person is asked whether a certain example is an instance of a specified property or relation.[5] Properties and relations of philosophical interest include knowledge, causation, reference, etc. Experimental philosophers of all stripes use survey methods to compare and contrast intuitions of groups of respondents. They imitate experimentalists in behavioral science by using controlled conditions of various sorts, e.g., slightly differently worded questions. Two types of stances, however, are taken toward the outcomes of these surveys (see Alexander, Mallon, and Weinberg, forthcoming). Members of the positive program of X-phi draw various kinds of inferences from differences in responses, but they don't fundamentally challenge the epistemologically soundness of intuition-based methods. Members of the negative program, by contrast, try to find evidence that raises doubts about the quality of intuitional evidence. Exactly how this challenge is mounted remains to be clarified. I shall interpret their findings as a challenge to the reliable-indicator properties of intuitions. But experimental philosophers themselves have not expressed the matter precisely this way. So it will take a bit of work to tease out the basis of the challenge. Even if my interpretation does not capture their intent, I think it is the most effective way to give skeptical punch to their line of thought.

The studies by negative experimental philosophers have generated a bevy of findings that they use to challenge the epistemic standing of classification intuitions. Several of these studies showed variation in intuitions across (American) subjects with different ethnic origins. Weinberg et al. (2001) gave their subjects a Gettier-like example featuring a protagonist named "Bob," and asked them whether Bob "really knows" a specified proposition or "only believes" it. A majority (74%) of subjects with Western ethnic origins responded that Bob only believes it, while a majority of subjects with East Asian (56%) and Indian (61%) ethnic origins responded that Bob knows. A different study by the same authors (Nichols et al., 2003) reported that the number of philosophy courses subjects had taken was a factor in influencing their responses. Subjects who had taken a number of philosophy courses were more susceptible to skeptical arguments than those who had taken fewer. A study by Machery et al. (2004) concerned intuitions about reference. They found that subjects with Western ethnic origins were more likely than those with East Asian ethnic origins to give causal-historical responses to queries describing Kripke's (1980) thought experiment involving the name 'Godel'. In another experiment, subjects' intuitions about Truetemp-like cases of putative knowledge were apparently influenced by the presentation of contrasting cases, in particular by the order of presentation of the various examples (Swain, Alexander, and Weinberg, 2008). Yet another finding in the literature is that intuitions are sensitive to affective content, such as seeing disgusting stimuli like greasy pieces of old pizza (Haidt, 2001; Nichols, 2004; Nichols and

Knobe, 2007; Prinz 2007). Subjects' intuitions about the relationship between causal determinism and moral responsibility can depend on the presence or absence of affective content in the narrative of the thought-experiment (Nichols and Knobe, 2007). Such data indicate that intuitions about the relationship between causal determinism and moral responsibility are "unstable".

Do findings of "instability" speak to the matter of intuitional unreliability? According to Alexander and Weinberg, these data show that intuitions are "sensitive to factors irrelevant to the content of the thought-experiments themselves" (2007: 61). In other words, people's intuitions are influenced by "irrelevant" factors. Sensitivity to irrelevant factors, argue Alexander and Weinberg, "impugns the status of intuitions as evidentiary" (2007: 66). How, exactly, does sensitivity to irrelevant factors bear on evidential status? Return to our suggested reliable-indicator construal of evidential status. To repeat an earlier example, a tree's having N rings on its trunk is evidence for its being N years old because a tree's being N years old is reliably correlated with its having N rings on its trunk. Now, if X is causally influenced by factors having nothing to do, correlation-wise, with the truth of Y, sensitivity to these factors will detract from the prospect that X is a reliable indicator of Y. For example, if things irrelevant to (i.e., uncorrelated with) a tree's age causally influence the growth of its rings, then the number of rings on its trunk won't reliably correlate with the tree's age. In general, findings that items of type X are influenced by factors irrelevant to Y undercut the thesis that items of type X are reliable indicators of Y. Hence, such findings would pose a challenge to the claim that items of type X are good evidence for facts of type Y.

The reliable-indicator construal of evidence also makes sense of the Alexander-Weinberg claim that order effects and affective-content effects challenge the evidential status of intuitions. If intuitions about knowledge are influenced by irrelevant factors such as whether the subject has just heard a contrasting case, those intuitions cannot be reliable indicators of the truth of the matter. Whether there was a prior presentation of another case is not diagnostic of the truth in the target case.

What about the inter-subject variation that experimental philosophers emphasize? How does this challenge the evidential status of intuitions? If a group of subjects performs a given thought experiment, and 70 percent make one intuitive judgment while the other 30 percent make a contrary judgment, they disagree with one another about the case. Assuming there is one correct answer, one group or another must be wrong: either 30 percent or 70 percent. Thus, at least 30 percent are wrong. If the breakdown is 50/50, then 50 percent must be wrong. So higher levels of variation imply higher (minimum) levels of unreliability. Under the reliable-indicatorship interpretation of evidence, variation poses a challenge to the evidential status of intuitive judgments.

Swain, Alexander and Weinberg (2008) provide ostensible confirmation that (putative) unreliability is the foundation of their challenge. They write:

> We certainly do not take ourselves to have offered anything like a general proof of the unreliability of all intuitions… (2008: 153)

This concession that they lack any proof of unreliability carries the implicature that unreliability worries are at the heart of their challenge.[6]

It is clear, then, how our interpretation of evidential status in terms of reliable-indicatorship makes sense of X-phi challenges to the evidential status of intuitions – although they themselves don't advance this interpretation. Moreover, our interpretation can proceed in terms of the framework we articulated earlier. Experimental philosophers should be understood to be presenting second-order evidence in support of the proposition that intuitions, or intuitive judgments, lack first-order evidential status. And I think we should concede at least this much to their claims: the cited experimental findings at least provide prima facie evidence in support of the denial of first-order evidential status. Of course, prima facie evidence is, by definition, defeasible evidence. So in admitting that there is some second-order evidence in favor of the non-evidential (or poor evidential) status of intuitions, we leave it open that this evidence can be defeated, or overridden. How might it be overridden? One route to defeat is an empirical route. The aforementioned experimental studies provide a body of empirical evidence that ostensibly conflicts with the claim that intuitions are reliable indicators -- for example, reliable indicators of the truth of their contents. But new empirical evidence might suggest a different conclusion. A second possible route to defeat is a non-empirical route. The X-phi arguments for intuitional unreliability crucially depend on (vague and tacit) assumptions about what the intuitions, or intuitive judgments, might be evidence for. There is room for philosophical maneuvering here. Maybe intuitions are highly reliable indicators for propositions that haven't yet been adequately considered. This is another possible way to sustain first-order evidential status for intuitions without disputing the existence of some second-order evidence pointing the other way.

There is also room for different grades of skepticism directed at intuitive judgments. A radical skepticism would claim that all intuitions are devoid of evidential value. A more qualified skepticism would only claim that some intuitions are devoid of evidential value. There are various ways that different intuitions might merit different evidential statuses. For example, intuitions experienced by some sort of experts might attain strong evidential status whereas intuitions experienced by other people might fail to attain such status. Or, the causal processes that generate some intuitions might be compatible with their outputs having high reliability whereas other causal processes might be incompatible with high reliability. (This comports with our earlier discussion of a link between reliable-indicatorship status and processes of causal production.) Alexander and Weinberg (2007) endorse a rather strong skeptical stance, adopting what they call a "restrictionist view" of intuitions' evidential status. Similarly, Machery et al. contend that "philosophers must radically revise their methodology" (2004: B9). More moderate views, however, would leave the door open to evidential respectability for some subset of intuitions or intuitive judgments. In the remainder of this paper, I shall try to pinpoint the kinds of additional work that should be done before clear conclusions can emerge.

6. A Framework for Studying Classification Judgments and Error Possibilities

Evidential respectability for intuitions, as for other possible sources of evidence, depends heavily on the choice of propositions for which the states or sources are taken to be evidence. When a detective claims that a certain item – say, a glove -- found at a crime scene is crucial evidence, he will have in mind some proposition or cluster of

propositions for which it might be evidence. These might include "Person X committed the crime," or "Person X was at the scene of the crime," etc. Similarly, when philosophers either claim or deny that intuitions are evidence, they usually mean that they are evidence for some proposition(s) of philosophical interest. However, I do not think there is complete agreement as to which proposition or propositions these are, exactly. This is open to debate. Part of the task of deciding whether classification intuitions have evidential value for philosophy is the task of deciding which hypotheses or conclusions (of philosophical interest) they might be evidence for.

One might think that the answer is straightforward. Isn't an intuition's content always the proposition for which it is, or purports to be, evidence? Not necessarily; there are other possibilities, as we shall see. Even if an intuition's propositional content is initially the most likely candidate, it is not the only one. Even if we assume that the reliability of a token intuition is determined by the truth of its content, there is a problem of using experimental data of the kind X-phi generates to assess the general reliability of intuitions (specifically, classification intuitions). Under our interpretation, experimental studies of the survey kind shed light on intuitional reliability by determining the amount of disagreement across subjects. But verbal disagreement may or may not reflect intuitive disagreement about one and the same proposition, because different subjects' intuitions may have different propositional contents (Nichols and Ulatowski, 2007; Sosa, 2008). This could emerge from divergent interpretations of the example they are asked to classify, of the target category they are asked to address, and so forth. Such complexities must be taken into account when trying to construct a general perspective on the kinds of evidence that might be relevant to determining the reliability of classification intuitions (and hence their evidential status).

I begin, then, with an analytical device: the "classification game" (cf. Ludwig, 2007). A classification game involves three players, the first two being the <u>interrogator</u> and the <u>respondent</u>. In the first move of a classification game the interrogator describes to the respondent an example or case, C, and a category, kind, property, predicate, or concept F, and asks whether C is a member or instance of F (or satisfies F). In other words, the respondent is invited to classify C -- or some person or event embedded in C -- as an F or a non-F. In the second move of the game, the respondent makes a spontaneous membership (or instantiation) determination: C belongs to F, or C does not belong to F. (Alternatively, he suspends judgment.) Finally, the respondent expresses this determination via a verbal report or mark on a questionnaire. The third player in the game is the <u>philosopher-scorekeeper</u>, who is assumed to be omniscient concerning what goes on in the heads of the interrogator and respondent, as well as other related matters. The scorekeeper's job is to ponder the question of whether the respondent's intuitive judgment – i.e., the mental state from which the verbal response issues – is true or false, and whether his making this judgment qualifies as a piece of evidence (for some hypothesis of interest). Assessing the judgment's truth-value may well require, of course, prior determination of its truth conditions, which is one of the philosophical questions on the table. To decide if it is a legitimate piece of evidence, the scorekeeper must also consider what it might be evidence <u>for</u>. Perhaps it is evidence for the proposition to which the respondent assents, e.g., that C is an instance of F. But what proposition is

that, exactly? Moreover, even if the respondent's intuition turns out to be poor evidence for that proposition, it might be good evidence for another proposition of interest to philosophy, as previously indicated.

I now introduce the technical term classifier to cover a wide range of things including properties, universals, kinds (natural or otherwise), linguistic meanings, and the contents of individuals' concepts. All these objects either have instances and counter-instances, or are standards or criteria that fix, pick out, or determine sets of instances and counter-instances. A central problem for the scorekeeper is to decide which type of classifier to use in keeping a respondent's (or group of respondents') reliability score. Such scores are readily influenced by the choice of classifier, as we shall soon see.

I divide classifiers into two categories. One category is person-specific or group-specific classifiers. A second category is free-floating classifiers. An example of the first category is the content of a single person's concept, a content the person associates with a predicate or term of interest 'F'. Assume that every (English-speaking) person associates a particular content with the term 'cup'. Any such content is a satisfaction condition that something must meet to qualify as a cup. Such a content supervenes (in part) on some complex mental state of the person, a state that, when activated, tends to gives rise to new thoughts and to verbal behavior, including classification behavior. The term 'concept' is used ambiguously to refer both to mental states of the requisite sort or to their contents. In the present article only the latter sort of thing is viewed as a classifier.[7] Obviously, one and the same content might be realized by the mental states of different individuals. Sometimes, however, we may want to refer to a particular person's concept (or conception) of a predicate 'F' (or F-ness), that is, to the content of that person's concept, ignoring whether or not that content is shared by others. In such a case we are interested in a person-specific classifier.

In addition to person-specific classifiers, we can take an interest in community-specific classifiers. For example, we might be interested in the meaning of 'cup' in English, that is, across the entire community of English speakers. Or we might take an interest in its meaning in a more restricted community of English speakers. Obviously, some words have different meanings in different sub-communities. In British English, 'boot' refers to the rear part of a car used for storage; in America 'boot' has no such meaning. I assume that meaning is a kind of content, a sub-propositional content. Thus, community-specific classifiers will also be contents. They presumably supervene, in some complex fashion, on language-related states and dispositions of the relevant community.

In addition to person-specific and community-specific classifiers, there may also be "free-floating" classifiers. These might include properties, universals, kinds, and so forth. These are not associated with, and presumably do not supervene on, any particular individual or group that represents them. Their nature or constitution does not essentially depend on their status as intentional objects of people's mental representations.

Classifiers in general are candidate targets of philosophical analysis or theorizing. When we try to analyze, explicate, or give a theory of knowledge, justice, causation, or personal identity, we aim to figure out what the satisfaction conditions are for 'knowing that p', 'being just,' and so forth. It is generally assumed that we are able to recognize instances of these classifiers. Our ultimate aim is to elucidate what those instances share. This consists of making explicit the "constitution" or "composition" of a selected classifier. Which type of classifier is appropriate? And for each candidate type of classifier, how do we go about investigating its constitution or composition? According to time-honored practice, we consider or think up actual and possible examples and ask whether they are, intuitively, positive or negative instances of the classifier. Which type of classifier makes best sense of this practice?

On any of these approaches to classifiers, there will presumably be genuine cases of knowing, causing, being the same person over time, etc. "in the world." Even the choice of person-specific classifiers as targets of philosophical investigation would not imply that causal relations or relations of personal identity over time are merely subjective rather than objective. Similarly, such a choice is fully compatible with there being genuine knowers in the world. However, it does not follow that the choice of classifiers is irrelevant to the viability of philosophical practice. To the contrary, a choice among classifiers can make a notable difference to the epistemic viability of philosophy's intuitional (or thought experimental) methodology, in particular, to its vulnerability to skeptical challenges.[8]

How can a choice of classifier affect error rates? Suppose that a person-specific content is our favored type of classifier, and that different people associate different contents with a term of philosophical interest. When invited to decide whether "knows" applies to a specified case, one respondent uses his concept of knowing, K_1, while another uses her concept of knowing, K_2. If the first makes a positive classification judgment and the second makes a negative one, their verbal responses suggest contradictory intuitions, so one of them must be in error. But if their judgments "answer" different questions, it does not follow that either is in error. Each might answer correctly in terms of his/her own concept, or conception, of knowing. Thus, choosing person-specific contents as the classifier would lower the error rate. This would result in higher reliability scores for intuitions and possibly avert intuition skepticism.

But, it will be objected, the propositions to which respondents assent are surely not propositions about the contents of their concepts. For example, if the posed query is whether a certain institution (e.g., slavery) is just, a respondent's answer presumably addresses justice itself, not his personal concept thereof. In saying that the specified policy is unjust, he means to classify it relative to a free-floating classifier, not relative to a person-specific or community-specific classifier.

Let P (below) be the evidential fact that consists in respondent R having an intuition with the content "Case C is an instance of justice." P* is the proposition that case C is an instance of justice.

P: Respondent R intuits that case C is an instance of justice.
P*: Case C is an instance of justice.

What might P be evidence for? One possibility, of course, is (the truth of) P*. This evidential relation between P and P* is the one people usually have in mind when they say that intuitions are evidence. But attitudinal facts can be evidence for matters other than their own contents. Thus, P might equally be evidence for, say, P**:

P**: R's concept of justice, J_1, is such that C is an instance of (or is in the extension of) the content of J_1.

An evidential link between P and P** has much to recommend it. Whatever R's understanding of justice – whether it accords with the moral structure of the universe or not – it is eminently plausible that R's intuition accords with (the content of) his own concept of justice. In other words, R would plausibly apply his own concept in ways that conform to that concept's content, or satisfaction conditions. Thus, if R assents to the proposition that case C is an instance of justice, isn't this evidence that C satisfies the conditions of his concept of justice, K_1, just as P** asserts?

Most philosophers are unimpressed by this kind of evidential "power" of intuitive judgments. They would prefer a tighter parallel between intuition and perception. With respect to perception, it is commonly held that propositions (facts) of the form Q are evidence for propositions of the form Q*:

Q: Person S seems to see that X.
Q*: X.

Similarly, it is possible that proposition P is evidence for P*. But that is not the evidential connection offered above. The P/P* evidential link is under threat from findings in experimental philosophy. At the moment, then, I consider the alternative evidential link between P and P**. Unlike the P/P* evidential link, the P/P** link is not undermined by experimental evidence. Moreover, if the philosopher/scorekeeper's job is to be on the alert for what states of affairs intuitions might reliably indicate, the P/P** link looks extremely promising.[9]

This sort of evidential scenario might motivate us to turn to person-specific contents as the favored target of philosophical analysis (cf. Goldman and Pust, 1998).[10] Alternatively, person-specific contents might be regarded as only a first target of philosophical analysis, on whose basis other targets might later be addressed (Goldman, 2007). I do not mean to preclude the latter scenario. However, most philosophers prefer some target of philosophical analysis that is not person-specific. So it is time to examine prospects for reliable evidence under the remaining types of candidate classifiers.

7. Problems for Evidential Status under Free-Floating Classifiers

The next class of classifiers I shall consider as candidates for the principal targets of philosophical analysis is free-floating classifiers (FFCs). I turn straightaway to a first worry, of ancient vintage. Under any approach to the proper type of classifier, respondents engaged in classification tasks will need to execute two prior tasks to be successful (i.e., accurate): (1) acquire a proper understanding of the case before them, and (2) acquire a proper grasp of the classifier in question. Since our focus is on the classifiers, set aside any potential problems with case understandings. The big question is how to acquire a grasp of FFCs. If such a grasp is not routinely attainable by ordinary people, reliable classification will be out of reach. Any defense of intuitional methodology by appeal to FFCs, then, owes us a story about how such a grasp can be routinely attained. Moreover, it must be attained for all of the specific classifiers to which intuitional methodology is customarily applied: knowledge, causation, object composition, personal identity, etc.

Divide the possible strategies for resolving this issue into two approaches: rationalist and empiricist approaches. A rationalist approach holds that people have a rational faculty of intellectual apprehension by which they grasp a priori the composition or constitution of abstract objects like universals. One finds this in Plato's treatment of the contemplation of the Forms and in much subsequent writing, of course. But what is rational apprehension, and how does it enable us to reliably grasp truths about the constitution of abstract objects? A majority of writing on this subject uses perceptual metaphors to characterize intellectual apprehension. It is said to be a way of "seeing" that such-and-such a proposition is necessarily true. But talk of "seeing" is clearly metaphorical, and does nothing to dispel the mystery. In contemporary science, perceptual processes like seeing and hearing are information-transmissional processes, which run from distal stimuli to mental experiences. Such processes are moderately well understood as sequences of causal links involving energy propagation, transduction, and transformation. In the case of genuine perceptual modalities, it is understood, at least in general terms, how information can be reliably transmitted from external objects to internal cognitive states (Dretske, 1981). But it remains a mystery how such transmission could work in the case of intuition, where the alleged medium is unknown to science. This makes naturalists suspicious of whether there could be any reliable process by which the properties of FFCs, i.e., their specific compositions, might be discerned.[11]

A second and less familiar puzzle about the rationalist solution is that it makes nonsense of customary philosophical practice. Customary philosophical practice, as we have emphasized, tries to get a handle on knowledge, causation, and so forth by working from singular cases in which knowledge, causation, etc. are allegedly involved. We "sneak up" on knowledge itself by making inferences from instances and counter-instances. Why do we proceed this way? Why approach the constitution of the classifier itself by eliciting intuitions about classifier instances? The reason must be, in part, because we lack any direct access to the classifiers themselves (and their constitutions). But if so, what should we make of the foregoing rationalist story? If we did have "direct" intellectual apprehension of the constitution of classifiers, why would we resort to a cumbersome, round-about, back-door method of soliciting intuitions about cases? Why not just call on our intellectual faculty to grasp the classifiers themselves directly, rather

than plod through the tedious process of devising hypothetical examples, recording intuitive responses to them, and drawing highly fallible explanatory inferences to the composition of the classifiers?[12]

Turning to the empiricist approach, perhaps respondents are thought to gain access to FFCs by empirical methods, including the confirmation of empirical theories. This is certainly the view of Kornblith (2002). But Kornblith does not seek to defend standard intuitional methodology. Our question is how FFCs (possibly including natural kinds) can be invoked as part of customary practice when the alleged access to them is via empirical theorizing. First, it's unclear what it would be like to have an empirical theory for many FFCs to which intuitional methodology could be applied. What would an empirical theory of object-composition (in the mereological sense) consist in? Second, innumerable empirical theories advanced during the course of history (whether scientific or non-scientific) have been false. Many (perhaps most) scientifically confirmed and accepted theories later proved to be false. So why should philosophers place much trust in people's theories of FFCs? Notice that philosophers routinely appeal to folk intuitions when using customary philosophical practice. Why should they assume that the folk have an accurate grasp of the nature of object-composition? If they lack such a grasp, though, how will their classifications of hypothetical examples get matters right? To repeat the critical premise, without accurate understandings of <u>both</u> hypothetical cases <u>and</u> the nature or constitution of a target classifier, people's judgments about whether the former instantiate the latter cannot be relied upon for accuracy. Yet philosophers trust such judgments to be reliable. No plausible rationale for such trust has been presented under the FFC approach.

8. <u>Community-Specific Classifiers</u>

Given the bleak prospects facing the FFC approach, let us turn to the remaining option: community-specific classifiers. In particular, consider linguistic meanings. As previously suggested, linguistic meanings are presumably fixed in part by the contents of person-specific contents that language users associate with various terms or predicates. These communal contents are not necessarily "averages" of person-specific contents. Language communities often assign extra weight, or deference, to experts. Thus, meanings might be thought of as "<u>normative"</u> contents that arise from individual concept contents. There is a problem here, however, because no individual user has direct access to this "social" entity (the communal meaning) in the way they have relatively direct access to their own individual concepts. Won't similar access problems arise for this type of classifier as for FFCs?

The problems are not as severe, however, as in the case of FFCs. Speakers acquire their personal concepts of most linguistic terms by interaction with others who are relatively authoritative speakers. Parents and caregivers are competent at identifying members and non-members of various terms' extensions, and learners receive feedback on their own classification attempts. Although such procedures do not infallibly transmit communal meanings, they are credible routes of access, far less obscure than those postulated for free-floating classifiers. More work in psycholinguistics and

developmental psychology is needed before the transmission of semantic competence to children is fully understood. But no comparable mysteries are encountered here as when we consider the prospects for FFCs. Therefore, I provisionally adopt this type of classifier as my favored type of classifier.

What, then, is the central type of hypothesis for which an intuition is evidence? The prototypical form of such a hypothesis will be, "Case C meets the satisfaction conditions for F," where F is the communal meaning of the queried predicate. This form of hypothesis is also the content of the subject's intuition; it is what the person intuits. I previously argued that intuitional states can be evidence for propositions other than (the truth of) their contents. But their propositional contents also remain good candidates to be hypotheses for which they are evidence. That is how we shall construe their main evidential significance for present purposes. Thus, if classification intuitions survive challenges to their evidential status, what they will be evidence for (in the first instance) is their contents.

Suppose, then, that a reader of the Gettier article intuits, with respect its first example, that Smith does not know that the man who will get the job has ten coins in his pocket (= T). This intuition might be evidence for the hypothesis that Smith fails to meet the satisfaction conditions of knowing with respect to T (according to the appropriate meaning of 'know' in the contextually relevant language community). To arrive at this intuitive udgment reliably, the intuiter presumably must make cognitive "contact" with a stored mental representation of the relevant sense of 'know'. He must (tacitly) activate that representation and compare its content with what he understands about T. In addition, his representation of the relevant sense of 'know' must correspond – closely enough – to a sense of the term 'know' in English. But none of these feats requires any miracles. If he has been suitably exposed to English and has normally operative language-acquisition capacities, he should have managed to acquire a relevant representation of 'know'. So it should not be a super-human task to be poised to have reliable intuitions about the proper classification of examples about knowing.

9. Processes of Classification Judgement: Preparatory and Final

Given the choice of communal meanings as our provisional type of classifier, let us now turn to a more detailed and wide-ranging exploration of the ways that empirical scientific investigation can and should be used to help assess the reliability of classification intuitions, and hence their evidential status. We have already seen how experimental philosophy has contributed to this exploration. But the familiar paradigms of experimental philosophy provide only hints about the processes that produce classification judgments. Identifying such processes, as we saw in section 4, can be extremely relevant evidence -- second-order evidence -- in assessing the evidential status of their end products. The standard X-phi studies provide only hints because survey studies are not designed to get at causal processes. Other types of psychological investigation, however, can and should fill this gap. Indeed, X-phi practitioners themselves sometimes emphasize the need for scientific studies outside the survey paradigm (Alexander, Mallon, and Weinberg, forthcoming).

Identifying the processes that influence classification judgments is a complex matter because multiple stages of processing are involved. Some processes are preparatory to classification proper and others are constitutive of it. Nonetheless, all can influence the final output. Given the complexity of these matters, we should not, of course, expect anything like precise truth-ratios to emerge from our discussion. Our aim is just to test the waters, to provide a preliminary sense of the sorts of processes that might exacerbate or moderate the threat to the evidential standing of intuitional methodology.

Working within the communal meanings framework, consider the cognitive tasks confronting a respondent in a classification game. She must first decode the question posed by the interrogator. This consists in decoding the case description and the target predicate. If predicate 'F' (or the principal lexical constituent of 'F') has multiple senses, a respondent must deploy semantic and/or pragmatic processes to identify the appropriate sense. There is room for respondent error here; and if error occurs at this stage, it magnifies the likelihood of error in the eventual classificational output.

Although errors can easily arise from insensitivity to the multivocality or polysemous character of many philosophical terms, I shall not pursue this source of error. Instead I shall consider a type of error that arises from the application of a certain polysemous term of great philosophical interest, viz., 'know'. The conventional wisdom in epistemology is that 'know' (in its propositional construction) has just one sense, the one that epistemologiy has doggedly pursued since Plato. In contrast, I believe there are at least three senses of 'know'. The first is mentioned with some frequency, though many epistemologists pooh-pooh it. This is the sense in which knowing something consists in being <u>completely confident</u> of it. A second sense of 'know' is to <u>believe something truly</u>. Having argued for this second sense in previous publications, here I just add (in the attached footnote) a new argumentative wrinkle.[13] The third sense of 'know' is the main quarry of epistemology, the <u>justified-true-belief-plus</u> sense of knowing (where 'plus' is a placeholder for an anti-Gettierization condition). For present purposes, I focus on the complete-confidence sense of knowing.

In the wake of discussions of contextualism, recent epistemology has lavished attention on ways that subtly different descriptions of one and the same case can prompt different responses from readers or hearers. Assorted explanations have been offered of these differences, including psychological explanations. For example, Hawthorne (2004) appeals to the so-called "availability heuristic" (posited by psychologists) to explain our unwillingness to impute knowledge when certain non-actualized possibilities of error are mentioned. Nagel (2009) offers an alternative explanation, which is interesting because the process she invokes is known from psychological work to be error-conducive. If this process shapes knowledge ascriptions and withholdings, it could nicely illustrate a preparatory process that encourages inaccurate inputs to classification.

Nagel first considers a "plain" story about John Doe in a furniture store, who is described as seeing a red table and believing that it's red. This plain story is contrasted

with an embellished version, which adds the information that a white table under red lighting would look just the same and Doe hasn't checked the lighting. Hearing the embellished story, says Nagel, is more likely to lead one to deny that Doe knows. Why? What transpires in a hearer's head to prompt this difference? Nagel suggests that hearing the second story may lead one to misrepresent Doe's mental condition by ascribing to Doe the same information about an error-possibility that the hearer herself has. This pattern of ascription would conform to a "bias" that psychologists call "the curse of knowledge." In the present case, the bias would manifest itself in a hearer first being herself focused on the possibility of deception through unusual illumination and then "projecting" this focus onto Doe. Why would projecting this thought onto Doe prompt a denial of knowledge to Doe? On Nagel's telling, once the hearer thinks about Doe as cognizant of the deception possibility, she will view him as overly hasty (hence unjustified) in believing the table to be red. I offer a slightly different explanation. Once the hearer represents Doe as being cognizant of the deception possibility, she will also represent him as being less than fully confident that the table is actually red. Then, thinking of knowledge as requiring complete confidence, she would deny that Doe knows. Plausibly, this denial of knowledge is a false classification judgment. Moreover, the process that leads to it is unreliable because it has an inherently error-generating feature, viz., attributing one's own informational states to a target.

It is instructive to note that the process of projecting one's own states onto others is also called the simulation process of mindreading. Elsewhere I defend the thesis that simulation is a primary method of executing mindreading tasks, and a good chunk of my evidence is precisely the documented error propensities that Nagel discusses (see Goldman 2006: 164-173). Interestingly, another theorist of philosophical methodology, Williamson (2007), takes simulation to be the method of making counterfactual judgments and hence the method of doing philosophical thought experiments. Williamson, however, seems unaware of the psychological literature on egocentric bias and its close association with simulation (at least in the case of mindreading). This is important because it poses the threat of error in precisely the domain he wishes to safeguard from error. He defends the epistemic quality of thought experiments on the ground that counterfactual judgments cannot be broadly inaccurate. But the very method he advances as crucial to counterfactual judgments is an error-prone method -- at least when applied to mindreading.

The error Nagel hypothetically pinpoints in a hearer's process of classifying Doe as a non-knower does not occur in the classification process proper. It occurs in the "run-up" to classification, when the hearer constructs in imagination what Doe is thinking. This illustrates what I call a "preparatory process," and it's intended to show how such processes can contribute to classification errors. In addition to the prospect of error-generating properties in preparatory processes, can a classification process proper also breed error on its own? Might this provide additional examples of how cognitive science can contribute second-order evidence that is relevant to the first-order evidential status of classification judgments? The answer, I think, is "yes." I offer two illustrations of this idea.

A bit of background is needed about the psychological structures underlying concept possession and how they are deployed during categorization. "Categorization" is the standard term in the psychological literature for classification proper. The psychological literature is full of such theories (for overviews, see Margolis and Lawrence, 1999; Machery, 2009). I shall not review all such theories, but shall simply select a couple of them for illustrative purposes.

The exemplar theory holds that concept possession consists of storing in memory a set of previously encountered exemplars of the category. Possessing a concept of "dog," for example, consists of storing in memory an assemblage of representations of dog exemplars. To categorize an object as a dog or non-dog, one (1) retrieves from memory some subset of exemplars from the total set and (2) compares them to the target for similarity. If the similarity is extensive enough, the target is categorized as a dog. If the target more closely resembles exemplars from a contrast category (e.g., "cat") than those of the dog category, it is categorized as a non-dog.

Suppose that the exemplar theory is correct. Focus on the fact that a selection is made from the exemplar-set during the retrieval process. What guides this selection? Perhaps contextual factors "prime" certain exemplars rather than others when the target is called to the subject's attention. This is the view of Medin and Schaffer (1978), whose "context model" of categorization assigns importance to context. This raises an important question. What fixes the content of an individual's personal concept under the exemplar theory? Is it fixed by all exemplars in the subject's memory and their properties? Or only by those exemplars retrieved from memory on a given occasion? If the latter were right, the concept's extension would be highly variable; it would vary in a manner that accords poorly with our customary conception of concept-content. Although an individual's concept can undergo some change as a result of new learning about a category, it shouldn't change as a function of the exemplars freshly sampled on each occasion. If we therefore make the natural assumption that a concept's extension should be fixed by all exemplars stored in memory, categorization might turn out to be highly fallible. Whenever a skewed choice of exemplars is used for comparison with the target, incorrect categorizations could easily be made.

This may well be what transpired in the experimental finding of order effects in knowledge attributions (Swain et al., 2008). Subjects were asked to indicate how firmly they agreed or disagreed with the statement that knowledge is achieved in a Truetemp-like case. Subjects were asked to make this rating either before or after rating other cases involving knowledge or non-knowledge. Subjects first presented with a clear case of knowledge were significantly less willing to attribute knowledge in the Truetemp case, whereas subjects first presented with a clear case of non-knowledge were significantly more willing to attribute knowledge in the Truetemp case. Plausibly, those first presented with a clear case of knowledge were, in effect, primed to retrieve a class of knowledge exemplars that set a high standard of comparison -- which induced less willingness to attribute knowledge in the ensuing Truetemp case. Those first presented with a clear case of non-knowledge were primed to retrieve a class of knowledge exemplars that set a much lower standard of comparison -- which induced greater

willingness to attribute knowledge in the ensuing Truetemp case. Assuming that at least one of these groups of attributions was rife with error, the mistakes seem attributable to the categorization process itself, which (by hypothesis) allows exemplar selection to play a major role in decision-making.

Another approach to concepts, the so-called theory-theory, takes exception to the idea that either exemplars or prototypes contain all of the information, or the most important information, stored in our concepts. Murphy and Medin (1985), Smith, Medin, and Rips (1984), Gelman (2003), and others argue for a "dual" or "hybrid" theory of concepts. This approach holds that concepts often store two kinds of information about their corresponding categories: an "identification procedure" and a "conceptual core". The first kind of information involves relatively superficial -- typically perceptual -- properties of previously encountered members of the category. The second contains information about deep, hidden, or essential features of members of the category. This is called the conceptual "core." Information contained in the conceptual core is often considered by subjects to be more diagnostic and central to the identity of the category than information contained in the identification procedure. Nonetheless, when quick categorization judgments must be made, and subjects lack detailed information about "deeper" properties of the instances to be categorized, they will use properties contained in the "identification procedure". Naturally, this can lead to errors.[14]

Thus, inspection of empirically based theories of categorization suggests that infallibility of judgment is not to be expected. It is therefore perfectly appropriate to worry about the level of reliability of categorization. This process cannot be assumed, a priori, to have a high enough reliability level (whatever "high enough" amounts to) to escape skeptical challenge.

10. A Dispositional Conception of Classificational Competence

Some philosophers might argue that a proper understanding of concept possession would avert the worries posed by X-phi and psychological theories of concepts and categorization. On this view, we can determine from the armchair that people are competent in deploying their own concepts. Classification dispositions are constitutive of the contents of their own concepts. So, as long as the scoring criterion features the contents of person-specific concepts, it will be a priori that people are competent and hence reliable. If so, empirical evidence is irrelevant, contrary to what philosophical naturalists contend. Concept possession, on this view, consists in having dispositions to sort or classify things. It's either a necessary or an a priori truth that if someone is disposed to classify certain things as Fs and others as non-Fs, then all the things he would classify as F are Fs (according to his concept of F, at any rate) and all the things he would classify as non-Fs are non-Fs (according to his concept). Hence, there is no room for error -- as long as we stick to the F-concept (or F-conception) of the individual doing the classifying.

Roughly this view is defended by Ludwig (2007).[15] Chalmers and Jackson (2001) spell out a similar idea in the following passages.

> If a subject possesses a concept and has unimpaired rational processes, then sufficient empirical information about the actual world puts a subject in a position to identify the concept's extension. For example, if a subject possesses the concept 'water', then sufficient information about the distribution, behavior, and appearance of clusters of H_2O molecules enables the subject to know that water is H_2O, to know where water is and is not, and so on. This conditional knowledge requires only possession of the concept and rational reflection, and so requires no further a posteriori knowledge. (2001: 323)

> If something like this is right, then possession of a concept such as 'knowledge' or 'water' bestows a conditional ability to identify the concept's extension under a hypothetical epistemic possibility, given sufficient information about that epistemic possibility and sufficient reasoning. That is, possession of these concepts in a sufficiently rational subject bestows an ability to evaluate certain conditionals of the form E → C, where E contains sufficient information about an epistemic possibility and where C is a statement using the concept and characterizing its extension, for arbitrary epistemic possibilities. And conceptual analysis often proceeds precisely by evaluating conditionals like these. (2001: 324)

It is not necessary to dispute the letter of these claims to draw a rather different moral. It may be true that in a sufficiently rational subject, sufficient information about a possibility bestows a conditional ability to identify the concept's extension. Nonetheless, it is doubtful that live subjects satisfy such idealized conditions. Human subjects may never be "sufficiently rational" and the information provided in thought experiments may characteristically be insufficient. In the terrain we are dealing with, and given our question about de facto reliability, it is not germane to appeal to the ideal operations of some ideal creature. Instead we should consider the customary operations of human classificational resources, warts and all. Chalmers and Jackson speak abstractly of concept possession, with no manifest concern for the actual routines or heuristics used in representing categories and making categorization decisions. We need to get "down and dirty" to see how human concept possession and categorization actually work to get a true measure of ways that they might go wrong. This, of course, invites us into the psychologist's laboratory -- or at least invites us to read the journals where laboratory work is reported and interpreted.

11. A Social-Epistemological Perspective on Intuitional Evidence

I now want to introduce a fresh element into the debate over the evidential value of intuitions. Suppose we assemble all of the sorts of considerations discussed to this point and decide that classification intuitions have a reliability of, say, .60, or .70. In other words, they are reliable indicators of the truth of their contents at roughly the .60-.70 level – greater than .50, but not terribly high. Would this be a defeat of intuition skepticism or a vindication? It could be argued either way. On the vindication side it could be argued as follows: "With that modest level of reliability, one surely isn't justified in believing (the truth of) an intuition's content"

(e.g., that Smith in the Gettier example doesn't know). "This modest level of evidential support isn't sufficient for flat-out belief." On the other hand, it could be argued that this level of reliability provides enough evidential support to entitle someone to adopt a moderately strong degree of credence in the intuition's content. That entitlement, or justification, constitutes a defeat of the skeptic – perhaps not a ringing defeat, but still a defeat.

Let us sidestep the potentially irresoluble question of what skepticism-avoidance requires (justifiedness of full belief or only something weaker). Let us simply ask whether there is any way to enhance the evidential force of intuition. There is such a way, I suggest, even without seeking new experimental findings or fresh intuitions. Until now our entire focus has been the evidential value of a single person's intuitions. But why should this question absorb all of our attention? We should be prepared to adopt a social-epistemological perspective and enjoy its potential fruits. It is a datum of philosophical practice (with clear parallels in other fields) that when a method yields a high degree of consensus among practitioners, or a highly replicable finding, levels of confidence rise. In the present case, the philosophical community is swayed when an intuition is widely shared about a particular case. When this occurred with the Gettier examples, the community's reaction was a strong conviction that the cases are not instances of knowledge. Are philosophers entitled to such a strong conviction? If a single person's intuitive judgment has only modest evidential weight, could it transpire that many people's intuitive judgments collectively have massive evidential weight?

Absolutely, and this is easy to explain in terms of the reliable-indicator approach to evidence conjoined with probabilistic facts. To say that one person's intuitions are reliable indicators of (say) the truth of their contents is to say that this person's intuitions have a probability of being right greater than 1/2. Suppose that all people whose intuitions are consulted are, in this sense, minimally competent. According to a well-known theorem, the Cordorcet Jury Theorem, the probability that a concurring majority of such competent individuals will be right is a swiftly increasing function of their total number, converging to 1 as their number tends to infinity (List and Goodin, 2001). In other words, as the size of the group increases, the reliability of a majority of the group being right – even a slight majority – increases dramatically. Even without unanimity, a majority of concurring individuals can have a collective probability of being right that greatly exceeds the probability of any one of them being right. Given the reliable-indicator approach to evidence, the weight of their evidence could therefore considerably exceed the evidential weight of a single individual's intuition. The theorem holds when three conditions are met: (1) there are two options, (2) each group member has a competence greater than ½, and (3) the members' beliefs are conditionally independent of one another. The independence condition is not trivially satisfied. If there are common causes of different individuals' beliefs, the independence condition isn't met. (Interesting results might still obtain even for interdependent beliefs, but such results are weaker than under the independence condition. See Dietrich and List 2004 for discussion). What may be inferred from this is that it is safe to trust concurring intuitions of a large group in a two-option choice if their intuitions are independent, even if the members are singly only somewhat competent. Whether these two conditions are satisfied is an empirical question.

What does this social-epistemological analysis imply about the justification that accrues to philosophers from intuitional evidence in the most favorable cases? Here is one way a critic

might respond. What the foregoing approach emphasizes is the evidential weight arising from a collective intuitional fact, e.g., the fact there is a certain distribution of intuitions about a Gettier case among practitioners in the field, for example, 900 to 20. But although this is a collective fact, and although this fact is a very good indicator of the truth of the content of the majority's intuition, this collective fact is not a piece of evidence possessed by any individual. Each individual possesses only his own intuitional piece of evidence; he doesn't possess the collective pattern of intuitions. Now, the critic will continue, a person is only justified or unjustified in virtue of items of evidence he possesses. So despite its initial attractions, the foregoing approach doesn't really help.

The critic raises a good point. We have not said enough about evidence possession to enable us to draw conclusions about justification. However, that matter requires our attention anyway (as acknowledged in footnote 1), so let us turn to it forthwith. Given that p is an item of evidence (which entails that p is true), under what circumstances is p evidence that person S possesses? Two principles of evidence possession may be proposed, one for basic evidence possession and the other for non-basic evidence possession.

(1) [Basic evidence-possession] If S is in a conscious mental state at t that is an evidential state, then S possesses that evidence at t.
(2) [Non-basic evidence-possession] If proposition p is true and S justifiably believes p at t, then S possesses p as an item of evidence at t.

Let us apply these principles to the question before us. If having an intuition is a piece of evidence, then according to principle (1) a person's having such an intuition implies his possessing that item of evidence. Obviously, this principle does not cover the case of a collective intuitional fact. No individual has (or is the subject of) such a collective intuitional fact. But nothing precludes an individual from justifiably believing that there is such a collective intuitional fact. He might have such a justified belief by receiving appropriate testimony from the 920 individuals in question. If so, then according to principle (2) he non-basically possesses this item of evidence, and his possession of this evidence might make him justified in holding a very high degree of belief -- or a categorical belief -- in the content of the widely shared intuition.

The foregoing principles only address the matter of evidence possession. They are silent on the question of whether an evidence possessor is justified in believing (or holding another attitude) toward a hypothesis that the (possessed) item of evidence supports. This is tricky terrain, and I won't try to settle matters there. It is tempting to require not only that the evidence possessed does support the hypothesis in question but also that the evidence possessor is justified in believing that the evidence supports that hypothesis. But this additional condition might be too strong. Specifically, it might be too strong for the case of basic evidence.

Finally, the preceding discussion suggests that the empirical study of intuitions and their processes of generation might be fruitfully re-directed. Instead of focusing on the traditional question of whether intuitions are "good enough" to qualify as evidence (or justification), we might instead ask how to structure or modify our philosophical classification games so as to maximize the evidential weight of intuitional outcomes. Both the choice of respondents and the

conditions in which their intuitions are elicited might be controlled and/or tweaked so as to avoid circumstances or situations that promote bias and error. Such improvements could be a constructive upshot of the originally negative program of experimental philosophy. It would yield an ameliorative approach to intuitional methodology, an approach that would facilitate the prospects of cognitive science playing a positive role in improving philosophical methods.

12. Intuitional Methodology and the A Priori

I promised at the outset to address two sorts of questions about intuitions or intuitional judgments. One is the question of whether such states constitute evidence at all for pertinent philosophical questions, and how strong such evidence is. The second is the question of what species of evidence intuitions provide: a priori evidence, a posteriori evidence, or some third type? In each case, of course, the importance of distinguishing between first-order questions and second-order questions was highlighted. I have now completed what I wish to say about the first question. In brief, without dismissing or minimizing the existing challenges to the evidential status of intuitions, there may still be reason to think that intuitions constitute bona fide evidence for philosophical conclusions of interest, at least in favorable cases (e.g., cases with high levels of intuitive agreement). To this extent our discussion is compatible with a traditionalist position that embraces intuitional methods in philosophy. At the same time, much of the second-order evidence needed to legitimize or secure intuitional methodology was said to be empirical, including evidence from empirical science. The latter thesis is definitely not traditionalist; it resonates with the program of philosophical naturalism. All of this discussion, however, concerned the legitimacy and strength of intuitional evidence, not the character or species of evidence it comprises. This final section turns to the taxonomical question: Is intuitional evidence a priori or a posteriori? And what type of second-order evidence is needed to address the taxonomical question?

It should first be noted that the terms a priori and a posteriori are not customarily invoked to classify types of evidence. Far more commonly they are used to classify types of knowledge or justification. So in this section I shall nudge the discussion over to the question of whether intuition-based beliefs have a priori justification or warrant.

The tradition draws the a priori/a posteriori distinction (whether applied to justification or to knowledge) by linking such classifications to the cognitive source of the justification or warrant. How shall we understand "source"? The straightforward understanding is in terms of the causal processes responsible for the warrant. A posteriori warrant is warrant produced by experiential, or perceptual, processes; a priori warrant is warrant produced by processes of intellection or ratiocination. Things get tricky, however, when we consider the possibility that a given justified belief might be the product of both experience and ratiocination. How should its justification be classified? The familiar answer accentuates the negative: a belief's warrant is a priori only if it is not influenced by any experiential elements, otherwise it is a posteriori. This is an unsatisfactory way to configure the taxonomy. It implies that warranted doxastic states with neither experiential nor ratiocinative sources are a priori. Belief based exclusively on memory, for example, has a source that is neither experiential nor ratiocinative. It is misguided, therefore, to classify its type of warrant as a priori.

A positive approach is therefore preferable, but how exactly should it be formulated? One approach is through a distinctive phenomenology. Plantinga (1993) pursues this line. He seeks to associate the a priori with a particular kind of intellectual "seeing," a non-perceptual, non-sensuous seeing. He is convinced that whenever one "sees" that a proposition a true, in this sense, there is a unique phenomenology involved. He despairs, however, of giving it a positive characterization. "Seeing", he says, is a way of forming a belief "with that peculiar sort of phenomenology with which we are well acquainted but which I can't describe in any way other than as the phenomenology that goes with seeing that such a proposition is true" (1993: 106).[16]

Philosophical naturalists will be skeptical of the claim that there is a distinctive phenomenology for all a priori beliefs. It is especially doubtful for classification beliefs, which is the (alleged) species of the a priori currently being examined. Even within the category of classification judgments, it is obscure what might be the common phenomenological thread among them. It is even more obscure, by my lights, what might be the distinctive phenomenological thread uniting all candidates for a priori status across the standard domains. If one weren't a rationalist philosopher with prior theoretical commitment to such a distinctive phenomenological unity, what are the chances that one would expect to find such a common thread across precisely these domains: mathematics, classification judgment, etc.? I regard the phenomenological unity thesis as a piece of highly "creative" speculation.

Many rationalists delineate the special character of their domain by reference to some modal properties of its distinctive propositional contents. Contemporary theorists of the a priori, however, have generally abandoned any simple link between a priori warrant and necessity. Fallibilism about the a priori is the dominant view these days, and rightly so (Casullo, 2003: chap. 3). What link between a prioricity and modal status, then, might be viable? Addressing the domain of classification judgment, Williamson (2007) proposes a tight link between knowledge of classification propositions and knowledge of associated counterfactuals -- a species of modal propositions. He interprets philosophical thought-experiments as episodes in which one contemplates a counterfactual proposition and renders a judgment on its truth-value. Does this point in the right direction for understanding a priori cognition -- at least in the domain of classification judgment?

Unfortunately, the proposed counterfactual interpretation of philosophical thought experiments is too strong, as Ichikawa and Jarvis persuasively argue (Ichikawa, 2009; Ichikawa and Jarvis, 2009). Applied to a Gettier counterexample, it demands that a counter-instance to the JTB analysis of knowledge be suitably close in modal space to make true the relevant counterfactual. This is unmotivated. Any possible instance of justified true belief without knowledge refutes the JTB account, whatever its location in modal space.

A second problem is how the a priori can be fitted into this picture. Williamson's off-line simulation account of counterfactuals assessment is not implausible, but it is questionable whether it qualifies as a priori. He himself raises doubt about this.

Furthermore, he rightly highlights important obscurities in the notion of the a priori. Specifically, it is difficult to know where to draw the line between experiential cognitive activities that play a purely evidential (justification-relevant) role in the run-up to a belief versus ones that play a merely enabling role. He comments:

> In such cases, the question "A priori or a posteriori?" is too crude to be of much epistemological use. The point is not that we cannot draw a line somewhere with traditional paradigms of the a priori on one side and the a posteriori on the other. Surely we can; the point is that doing so yields little insight. (2007: 169)

To the extent that this complaint expresses a general dissatisfaction with the a priori/a posteriori distinction, it resonates with a dissatisfaction of my own in an earlier article (Goldman, 1999a). My main dissatisfactions there were with the simplistic dichotomy of epistemic sources and its lack of attunement to the psychological literature. If one wants to draw fundamental distinctions between psychological sources of belief, there are many other, more promising paths to take. One might follow the modularity path, and divide sources of belief in terms of the mental modules from which they emanate. This might indeed pinpoint numerical cognition and logical cognition as interesting sources of belief. But they would also wind up distinguished from one another, not lumped together as the traditional taxonomy proposes. Another promising path is to adopt the recent "dual-process" model of the mind and seek to understand different epistemic statuses in terms the distinction between automatic versus controlled levels of processing (Evans and Frankish, 2009). Either program is likely to yield more fruitful insight into diverse cognitive methods or mechanisms than the antiquated distinction epistemology has saddled itself with for far too long. Perhaps the old distinction was worth retaining when it was thought to track a distinction between grades of justification. But now that it is pretty much universally conceded that a priori status need not correlate with superior strength or degree of warrant, why retain a distinction with so little to be said for it?

Nonetheless, traditionalists will remain anxious to retain the a priori category and insist that intuitional judgments exemplify the category. This move is critical if they are to preserve the sanctity of the armchair method for philosophy (where armchair method is equated with a priori method). Moreover, classification judgments, at first blush, are a clear example of a priori cognition. Don't they essentially involve determinations of conceptual relations between two contents? Confronted with a Gettier case, one (tacitly) accesses the content of one's knowledge concept and the content of the specified scenario and reasons to the conclusion that what satisfies the latter (with certain added presuppositions) does not satisfy the former. Isn't this an unexceptionable example of non-experiential and ratiocinative cognition?[17]

While this description may be a true and relevantly complete account of some processes of classification judgment, it doesn't fully describe other such processes. In particular, philosophers often consult with their students and peers about examples; they try out the scenarios on others and solicit their reactions. The testimonies they receive on the matter -- whether they support the subject's own initial view or provide defeaters for

it – are all justificationally relevant to the subject's view, either positively or negatively.[18] Since all such testimony arrives via perceptual means, these are a posteriori components of the total mental process that is evidentially relevant to the judgment's epistemic status. On the standard approach, then, the judgment should be a posteriori. Certainly it isn't purely a priori.[19]

Of course, the foregoing "incursion" of empirical elements into the domain of classification judgment is not an incursion involving empirical science. The mere sharing and comparing of intuitions, as done by ordinary philosophers of a non-experimental stripe, does not demonstrate the relevance of empirical science to the question of a priori status for intuitional judgments. Next, however, we turn evidence of this other kind. There is evidence from cognitive science that challenges the orthodox assumption that the cognitive processing of relations between concepts is a purely non-experiential, i.e., "non-modal" activity (in the psychologist's sense of "modal"). The orthodox view of concepts in both philosophy and cognitive science regards them as "abstract" or "amodal" mental representations, that is, representations devoid of perceptual contents. But a growing wing of cognitive science led by Barsalou (1999, 2003, 2009) has accumulated some fairly striking empirical evidence suggesting that conceptual processing heavily recruits modality-specific (i.e., perception-derived) materials. In philosophy this neo-empiricist theme has been developed and defended by Prinz (2002). While still a minority view, it shows that it cannot be taken for granted that conceptual processing is a non-experiential affair. This is a scientific question! (Nor can it be resolved by introspection. The level of processing that might include modality-specific representations is below the level of consciousness, hence inaccessible to introspection.)

Here is a brief sample of the evidence Barsalou marshals in support of his neo-empiricism (see Barsalou et al., 2003). In one set of experiments, participants were asked to list the features of a concept verbally. One group was instructed to use imagery during their performance whereas the others (the "neutral" group) were asked to use whatever representations they chose. The neutral group showed very substantial similarity to the imagery-instructed group, as judged by their response times and errors. Moreover, perceptual similarity has been found to affect property verification in conceptual tasks. When participants were asked to verify whether mane is a property possessed by pony, they were faster if mane had been previously verified for horse as contrasted with being verified for lion. Apparently, verifying mane for horse primed a perceptual shape for mane that was later used to verify mane for pony, whereas verifying mane for lion activated a different shape that could not be so used.

The role of modality-specific representations in conceptual processing is further confirmed by evidence from cognitive neuroscience. Studies by Martin and colleagues (Martin et al., 2000; Martin and Chao, 2001) have localized various forms of conceptual processing in modality-specific brain areas. Conceptualizing object motion, for example, activates areas near brain centers for perceiving motion; and conceptualizing actions that an agent could perform on an object activates the motor system (considered a modality-specific code as contrasted with an abstract code). All of this evidence gives suggestive empirical support to the notion that manipulating concepts -- including applying them to examples -- involves perception-related cognition, which clashes with the classical picture of non-experiential, a priori cognition.

I am raising doubts about the contention that classification judgments have an a priori status. But I am not insisting on this. In particular, I am prepared to admit that classification judgments have a partly a priori status, stemming from the ratiocinative character of the principal processes that influence classification judgment. Is this an odd concession from a self-proclaimed philosophical naturalist? Is it an abandonment of naturalistic scruples! Not so. As argued in (Goldman 1999a), a priori warrant need not be wholly anathema to naturalism. It is just a status that arises from a certain class of cognitive processes. However, in light of our earlier remarks about the "new" a priori -- which admits fallibility and lays no claim to extraordinary evidential strength -- naturalists have much less reason to decry or resist it. Naturalists are suspicious of the a priori partly in virtue of its historically "elitist" claims, especially its claim to infallibility. Since recent a priorists have substantially toned down their claims and their rhetoric, there is far less cause for resistance.

A final point. Even if cognitive scientific evidence were to push us in the orthodox direction, revealing that classification judgment is a purely amodal cognitive activity, this would not affect the "big picture" of philosophical methodology I highlighted from the start. As I stressed from the beginning, philosophical methodology "as a whole" includes two levels: (1) making classification judgments about cases and drawing inferences about the contents of target classifiers, and (2) engaging in second-order investigations intended to illuminate the reliability and character of activities at the first level. These second-order investigations are ones that the philosophical profession has an (epistemic) obligation to undertake. It must scrutinize the worthiness and appropriateness of its first-level methodology, and this should be done using all relevant tools at its disposal, including scientific tools. Members of any profession, especially ours, should seek justification for believing that conclusions reached via its dominant methodologies are justified conclusions, well supported by all relevant evidence. We should seek to be justified in believing that many or most of the philosophical judgments we make are justified. If this requires appeal to existing scientific evidence -- or the generation of fresh scientific evidence by means of new experimentation -- then so be it. These kinds of evidence must be consulted and/or gathered. Furthermore, I have argued that the justification of existing philosophical methods (as well as their improvement) does require such scientific evidence. These theses comport well with the outlook of philosophical naturalism.[20]

Bibliography

Alexander, J., Mallon, R. and Weinberg, J. M. (forthcoming). "Accentuate the Negative." Review of Philosophy and Psychology.

Alexander, J. and Weinberg, J. M. (2007). "Analytic Epistemology and Experimental Philosophy." Philosophy Compass 2(1): 56-80.

Barsalou, L. W. (1999). "Perceptual Symbol Systems." Behavioral and Brain Sciences 22: 577-660.

Barsalou, L. W. (2009). "Simulation, Situated Conceptualization, and Prediction." Philosophical Transactions of the Royal Society B 364: 1281-1289.

Barsalou, L. W., Simmons, W.K., Barbey, A.K., and Wilson, C.D. (2003). "Grounding Conceptual Knowledge in Modality-Specific Systems." Trends in Cognitive Sciences 7(2): 84-91.

Bealer, G. (1998). "Intuition and the Autonomy of Philosophy." In DePaul and Ramsey (eds.), Rethinking Intuition: The Psychology of Intuition and Its Role in Philosophical Inquiry. Lanham, MD: Rowman and Littlefield.

Blanchard, T. (2009). "Conceptual Analysis, Similarity-Based Accounts, and the Theory-Theory of Concepts." Seminar paper, Rutgers University.

Brown, J. (forthcoming). "Thought Experiments, Intuitions, and Philosophical Evidence." Dialectica.

Burge, T. (1979). "Individualism and the Mental," in Midwest Studies in Philosophy, 4: 73-121.

Burge, T. (1993). "Content Preservation," Philosophical Review 102: 457-488.

Carston, R. (2002). Thoughts and Utterances. Malden, MA: Blackwell.

Casullo, A. (2003). A Priori Justification. New York: Oxford University Press.

Chalmers, D. J. and Jackson, F. (2001). "Conceptual Analysis and Reductive Explanation," Philosophical Review 110(3): 315-360.

Christensen, D. and Kornblith, H. (1997). "Testimony, Memory, and the Limits of the A Priori," Philosophical Studies 86: 1-20.

Dietrich, F. and List, C. (2004). "A Model of Jury Decisions Where All Jurors Have the Same Evidence," Synthese 142: 175-202.

Dretske, F. I. (1981). Knowledge and the Flow of Information. Cambridge, MA: MIT Press.

Evans, J. St. B. T. and Frankish, K. (eds.) (2009). In Two Minds: Dual Processes and Beyond. Oxford: Oxford University Press.

Gelman, S. (2003). The Essential Child. New York: Oxford University Press.

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

Goldman, A. I. (1986). Epistemology and Cognition. Cambridge, MA: Harvard University Press.

Goldman, A. I. (1999a). "A Priori Warrant and Naturalistic Epistemology." In J. Tomberlin (ed.), Philosophical Perspectives 13: 1-28. Malden, MA: Blackwell. Reprinted in Goldman, A. I. (2002). Pathways to Knowledge: Public and Private. New York: Oxford University Press.

Goldman, A. I. (1999b). Knowledge in a Social World. Oxford: Oxford University Press.

Goldman, A. I. (2006). Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading. New York: Oxford University Press.

Goldman, A. I. (2007). "Philosophical Intuitions: Their Target, Their Source, and Their Epistemic Status." Grazer Philosophische Studien 74: 1-26.

Goldman, A. I. (forthcoming). "Epistemic Relativism and Reasonable Disagreement." In R. Feldman and T. Warfield (eds.), Disagreement. New York: Oxford University Press.

Goldman, A. I. and Olsson, E. (2009). "Reliabilism and the Value of Knowledge." In A. Haddock, A. Millar, and D. Pritchard (eds.). Epistemic Value. Oxford: Oxford University Press.

Goldman, A.I. and Pust, J. (1998). "Philosophical Theory and Intuitional Evidence." In M. DePaul and W. Ramsey (eds.), Rethinking Intuition: The Psychology of Intuition and Its Role in Philosophical Inquiry. Lanham, MD: Rowman and Littlefield.

Grundmann, T. (2010). "Some Hope for Intuitions: A Reply to Weinberg."

Haidt, J. (2001). "The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment," Psychological Review 108: 814-834.

Hawthorne, J. (2004). Knowledge and Lotteries. Oxford: Oxford University Press.

Ichikawa, J. (2009). "Knowing the Intuition and Knowing the Counterfactual. Philosophical Studies 145(3): 435-443.

Ichikawa, J. and Jarvis, (2009). "Thought-Experiment Intuitions and Truth in Fiction." Philosophical Studies 142(2): 221-246.

Kelly, T. (2008). "Evidence", Stanford Encyclopedia of Philosophy, E. N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2008/entries/evidence/>.

Kornblith, H. (2002). Knowledge and Its Place in Nature. Oxford University Press.

Kripke, S. (1980). Naming and Necessity. Cambridge, MA: Harvard University Press.

Lewis, D. K. (1983). "New Work for a Theory of Universals," Australasian Journal of Philosophy 61: 343-377.

List, C. and Goodin, R. (2001). "Epistemic Democracy: Generalizing the Condorcet Jury Theorem." Journal of Political Philosophy 9(3): 277-306.

Ludwig, K. (2007). "The Epistemology of Thought Experiments: First Person versus Third Person Approaches." In H. Wettstein (ed.), Midwest Studies in Philosophy 31.

Lycan, W. (1988). "Moral Facts and Moral Knowledge," Judgement and Justification. New York: Cambridge University Press.

Machery, E. (2009). Doing without Concepts. Oxford: Oxford University Press.

Machery, E., Mallon, R., Nichols, S. and Stich, S. (2004). "Semantics Cross-cultural Style." Cognition 92(3): B1-B12.

Margolis, E. and Lawrence, S. (eds.) (1999). Concepts. Cambridge, MA: MIT Press.

Martin, A. et al. (2000). "Category-specificity and the Brain: The Sensory-Motor Model of Semantic Representations of Objects." In M. Gazzaniga (ed.), The New Cognitive Neurosciences, 2nd edition, pp. 1023-1036. Cambridge, MA: MIT Press.

Martin, A. and Chou, L. (2001). "Semantic Memory and the Brain: Structure and Process." Current Opinion in Neurobiology 11: 194-201.

Medin, D. and Schaffer, M. (1978). "A Context Theory of Classification Learning." Psychological Review 85: 207-238.

Mnookin, J. (2008). "Of Black Boxes, Instruments, and Experts: Testing the Validity of Forensic Science". Episteme, A Journal of Social Epistemology 5(3): 344-358.

Murphy, G. and Medin, D. (1985). "The Role of Theories in Conceptual Coherence." Psychological Review 92: 289-316.

Nado, J. (submitted). "Intuition, Evidence, and the Threat of Skepticism." Department of Philosophy, Rutgers University.

Nagel, J. (2009). "Knowledge Ascriptions and the Psychological Consequences of Thinking about Error." Philosophical Quarterly.

Nichols, S. (2004). "After Objectivity: An Empirical Study of Moral Judgment." Philosophical Psychology 17(11): 3-26.

Nichols, S., Stich, S., and Weinberg, J. M. (2003). "Meta-Skepticism: Meditations in Ethno-Epistemology." In S. Luper (ed.), The Skeptics: Contemporary Essays. Aldershot: Ashgate.

Nichols, S. and Knobe, J. (2007). "Moral Responsibility and Determinism: The Cognitive Science of Folk Intuitions." Nous 41(4): 663-685.

Nichols, S. and Ulatowski, J. (2007). "Intuitions and Individual Differences: The Knobe-Effect Revisited," Mind and Language 22(4): 346-363.

Prinz., J. (2002). Furnishing the Mind. Cambridge, MA: MIT Press.

Prinz, J. (2007). "Can Moral Obligations Be Empirically Discovered?" In H. Wettstein (ed.), Midwest Studies in Philosophy, 31.

Smith, E., Medin, D. and Rips, L. (1984). "A Psychological Approach to Concepts: Comments on Rey's 'Concepts and Stereotypes'." Cognition 17: 265-274.

Sosa, E. (2007). A Virtue Epistemology, Apt Belief and Reflective Knowledge. Oxford: Oxford University Press.

Sosa, E. (2008). "Experimental Philosophy and Philosophical Intuition," in J. Knobe and S. Nichols, eds., Experimental Philosophy. Oxford: Oxford University Press.

Sperber, D. and Wilson, D. (1998). "The Mapping between the Mental and the Public Lexicon," in P. Carruthers and J. Boucher (eds.), Language and Thought. Cambridge: Cambridge University Press, pp. 184-200.

Swain, S., Alexander, J. and Weinberg, J. M. (2008). "The Instability of Philosophical Intuitions: Running Hot and Cold on Truetemp." Philosophy and Phenomenological Research 76(1): 138-155.

Weinberg, J. M. (2007). "How to Challenge Intuitions Empirically without Risking Skepticism," Midwest Studies in Philosophy 31(1): 318-343.

Weinberg, J. M., Nichols, S. and Stich, S.P. (2001). "Normativity and Epistemic Intuitions." Philosophical Topics 29: 429-460.

Williamson, T. (2007). The Philosophy of Philosophy. Malden, MA: Blackwell.

[1] It is unclear exactly which set of intuitions, or judgments, experimental philosophers mean to challenge with respect to evidential status. Do they mean to challenge the evidential status of classification judgments in general? That would constitute a very wide-ranging and radical skepticism. An alternative approach is to defend a more selective skepticism (see Nado, submitted), i.e., an epistemological challenge to some subset of classification intuitions or judgments. Which subset? First, it would definitely be restricted to classification judgments rendered in response to imagined, or verbally described, hypothetical cases. Second, it may be restricted to judgments concerning philosophical categories, not ordinary physical-object categories such as "table" or "cat". Third, it may be restricted to judgments concerning difficult philosophical cases. These are, after all, the cases that receive the bulk of discussion in the philosophical literature and are certainly the cases most intensively researched by experimental philosophers. They are also, of course, among the most important cases that require resolution for purposes of philosophical theory testing. Well-grounded doubts about the evidential status of any of these subsets would be of interest. Clearly, however, the wider the set, the more interesting the challenge. I shall not try to settle precisely which set of cases is to be targeted. I leave this factor unresolved, as it is left in most of the experimental philosophy literature. (For some attention to it, though, see Weinberg, 2007, especially pp. 334-335.)

[2] This terminology concerns "objective" states of evidence and evidential relations. But we must also be concerned with when an item of evidence is evidence for a given person. In other words, we shall want to be concerned with what it is for someone to possess evidence. This topic is postponed until section 11.

[3] The intuiting/intuited distinction is due to Lycan (1988), who also opts for intuiteds rather than intuitings as the primary evidential entities.

[4] A thorough-going rebuttal can be found in Brown (forthcoming).l

[5] Typically it will be a person or event featured in the example that will be the candidate bearer of the property in question, not the example (or scenario) as a whole. For convenience, however, I shall often speak as if the example is the candidate for instantiating or exemplifying the target property or relation.

[6] Weinberg (personal communication) comments that this was not his intention. Indeed, in Weinberg (2007) he offers a different account of the epistemological weakness of intuitional methods, namely, that their possible errors are not open to our detection and correction. However, even if the unreliability flaw is not Weinberg's primary complaint, it may have been what his.co-authors intended. At a minimum, it is a reasonable idea for others to advance. Concerning Weinberg's own view, see Grundmann (2010).

[7] In earlier writings (Goldman and Pust, 1998; Goldman, 2007) I have sometimes muddied the waters by talking about 'mentalist' and 'extra-mentalist' classifiers and treating concepts in the psychological sense (rather than the content sense) as one type of classifier, indeed, a type that should be favored as the best for philosophical purposes. Here I offer a cleaned-up view of the terrain, which descends from the earlier treatment but is intended to improve upon it.

[8] These considerations represent a different perspective than the one defended by Sosa (2007): "[W]e can engage in our … philosophical controversies, and regard them as objective, without ever raising the question of the ontological status of the entities involved, if any. Mostly we can conduct our controversies … just in terms of where the truth lies with regard to them, leaving aside questions of objectual ontology." (2007: 59) The position defended here, by contrast, holds that questions of objectual ontology cannot be ignored. Otherwise, we cannot be clear about which propositions intuitions purport to be right about, or provide evidence for, a matter that is crucial to intuitional reliability.

[9] This linkage will be particularly attractive if one adopts a strongly pragmatic approach to the interpretation of lexical items, especially the sort of approach associated with relevance theory (Sperber and Wilson, 1998; Carston, 2002). According to this approach, word meanings "do not express full-fledged

concepts, but rather concept schemas, or pointers to a conceptual space, on the basis of which, on every occasion of use, an actual concept (an ingredient of a thought) is pragmatically inferred" (Carston, 2002: 360). If this is right, there are few concepts stably expressed by public words or phrases. The content of an individual's thought on a particular occasion, and the truth-conditions it aspires to fulfill, is very much a function of pragmatic elaboration. Thus, the evidential reliability of intuitive judgments is likely to be maximized by considering the propositions a subject constructs "on the fly." These propositions are not fully fixed or adequately captured by public word meanings. I am not endorsing this position, but merely commenting on the evidential story one might reasonably prefer if one were to adopt it.

[10] Does our treatment of "personal concepts" tacitly amount to "individualism" about the contents of (sub-propositional) attitudes? Hasn't Burge (1979) blown individualism out of the water? How, then, can we take seriously the content of "person-specific" concepts or contents, which sounds suspiciously like the endangered species "narrow content"? I would reply that however strong is the case for some social component in determining the contents of propositional and sub-propositional attitudes, that component must be limited. Inter-individual differences must be honored. Even if the correct use of the verb 'refute' makes it an achievement term, more than a few English speakers nowadays use it in a different sense, i.e., (merely) try to prove that someone is mistaken. (Thanks to Bob Matthews for the example). There is a similar discrepancy between the correct use of 'infamous' and the widespread (mis-)use in which it is synonymous with "famous". We must clearly preserve space for differences in person-specific concepts and not allow an individual's idiolect to be swamped by proper usage.

[11] It might be suggested that, for some abstract entities, namely Lewis's (1983) "natural" properties, it isn't that we have a special epistemic power to grasp them but that they have a greater propensity, or "eligibility," to become the objects of our thoughts. This should not strike (science-oriented) philosophical naturalists as terribly helpful or illuminating. How does it transpire that certain privileged abstract objects insinuate themselves into our thoughts or thought contents? This is as objectionably mysterious as any supposed cognitive faculty for "grasping" such objects.

[12] One rationalist, Bealer, candidly admits that the most we can establish about our access to philosophically interesting classifiers is the possibility of grasping them fully (Bealer, 1998). But if all we can secure is the mere possibility of full access, why should we have confidence in the ability of our intuitions to track the truth about purported "matches" or "fits" between a case and a classifier?

[13] Sense (ii) is what I call a "weak" sense of 'know' (Goldman 1999b; Goldman and Olsson, 2009). To defend the existence of such a sense, consider the following. Lucy is reading a mystery. She knows that her friend has already read the book, so she says to him: "Don't tell me how it ends. I don't want to know how it ends, because that would ruin the story." What does she mean by "I don't want to know how it ends"? If the only sense of (propositional) knowing were given by (iii), the state she says she doesn't want to occupy would be the state of having a justified and unGettierized belief in a truth about the ending. Is this the state Lucy wants to avoid? To avoid this state, it would suffice to have an unjustified belief in some such truth, or to have a justified but Gettierized belief in such a truth. Surely, however, these are not states Lucy would be happy to occupy. She wants to avoid them too. What she means is that she doesn't want to believe any truth about the ending. So she seems to use 'know' in a sense that consists of believing a truth. This is the "weak" sense specified by (ii).

[14] My discussion of the theory-theory approach has benefited from Blanchard (2009), although his concerns are rather different from mine. Chalmers and Jackson may respond that in this case the subject lacks (ideally) sufficient information to make a categorization. True enough. But almost all thought experiments created by philosophers are underspecified in the sense that suitable additional specifications could affect the proper classification. That is even true of Gettier cases. So we can hardly insist that the reliability of classification processes be scored only by reference to cases in which subjects receive "ideally sufficient information" about the target case.

[15] I don't mean to imply that Ludwig's account is as simplistic as the one I present here. I introduce this mainly to provide a foil for the "structure-and-process" conception of concepts that I proceed to discuss in the next section. I myself suggested something vaguely similar to the dispositional account in Goldman (2007: 14-15)..

[16] Which kind of proposition does he mean in speaking of "such" a proposition? He cannot mean a necessary proposition because he admits that our a priori beliefs are fallible. Apparently, we can "see" that such a proposition is true and necessary when it isn't really necessary.

[17] Of course, in the normal play of a classification game, a respondent reads or hears (1) an interrogator's description of a case and (2) his query about whether the case exemplifies a specified classifier. So the respondent's cognitive episode begins with perceptual experience. This poses an obvious obstacle to the (complete) episode qualifying as justified a priori. However, we are obviously "subtracting" the initial segment of the cognitive episode. The part that interests us is the post-perceptual analysis of the relationship between the "case" and queried predicate. Since the case could have been invented by the respondent herself – without any major change in the "core" of the methodology -- I am setting aside the perceptual input aspect.

[18] I assume here that there is no inference from the testimony of others, at least at the time of the current judgment. Such inference would eliminate the status of the subject's belief as an intuition (since intuitions were defined as spontaneous, i.e., non-inferential, judgments). Still, the subject's intuitive judgment could have been influenced by previous testimony.

[19] Burge (1993) defends the view that beliefs formed on the basis of testimony are, in many cases, a priori warranted, arguing that the role of perception in testimony is not justificatory. However, I agree with Christensen and Kornblith (1997) that this view is implausible. As they say, in interlocution it is precisely perception that warrants one in taking a certain proposition to have been presented as true, and therefore plays a warranting role as in typical empirical beliefs (1997: 5).

[20] The final version of this paper has benefited from discussions of earlier versions presented in several venues, as well as with individuals outside of those venues. An early draft was read and discussed in my graduate seminar on philosophical naturalism at Rutgers in the fall of 2009. Many participants made valuable contributions in the context of that seminar, but I especially profited from criticisms, and suggestions by Robert Beddor, Thomas Blanchard, Jenny Nado, Lisa Miracchi, and Blake Roeber.. Another version was presented as the (official) Romanell Lecture at the Pacific Division meeting of the APA, where Jonathan Weinberg and Ron Mallon made constructive remarks during the discussion period. Subsequent revised versions were given at Arche Philosophical Research Centre at the University of St. Andrews and an epistemology conference at PUCRS in Porto Alegre, Brazil. Very helpful feedback was received in both cases. I also profited from hearing Jennifer Brown deliver her paper on this topic (cited here) at Arche. Finally, I have received extensive comments and pointers of great value from Thomas Grundmann and Holly Smith.