Everything you always wanted to know about knowledge but were afraid to ask

Gregory Wheeler
Centro de Inteligência Artificial (CENTRIA)
Departamento de Informática, Universidade Nova de Lisboa
2829-516 Caparica, Portugal
greg@di.fct.unl.pt

July 4, 2005

Abstract

This lecture was delivered to the Departamento de Informática, Universidade Nova de Lisboa in February 2004 and the Technical University at Dresden in July 2005. Please do not cite.

1 Introduction

Susan Haack has quipped that the surest way to clear the room at a cocktail party is to utter the word 'epistemology'. One goal of this course is to strip the pretense from this core topic of western philosophy and present what might be considered the nuts and bolts of the theory of knowledge. My goal will be to offer a guided tour of the field, putting up sign posts along the way that, first, highlight the main issues exercising the field today and, second, highlight the subset of these traditional issues that I think are interesting from a computational point of view. Hence, my interest here is to present the traditional distinctions and approaches to the subject and introduce you to the terms of the art so that the descriptions of our group's future contributions will be recognized as such by the appropriate communities. This is the broad agenda guiding this course.

1.1 Getting a grip on 'knowledge'

Like butchers to a cow, there is a traditional way philosophers have chopped this concept down to size. As a first pass, notice that there are distinct kinds of knowledge attribution. Tradition recognizes three distinct kinds of knowledge attribution, while some suggest that there are really only two fundamental kinds, with the stray notion subsumed as a proper class of one of the two main notions. This detail aside, all agree that in common language (English, here),

with respect to an epistemic agent (subject) S, the sentence 'S knows x' may be understood to attribute to S one of three things, namely

1. that S is *acquainted* with x, such as 'Greg knows Carlos Damásio', meaning that I am acquainted with Carlos. Compare this to 'Greg does not know George Bush'. There are several parameters to pin down to evaluate the truth of these propositions, but the point to notice here is that I know more facts (propositions) about George Bush than I do about Carlos Damásio, but I nevertheless am acquainted with Damásio but not acquainted with Bush. *Knowledge by acquaintance* is the name commonly given this first type of knowledge attribution. Some philosophers have proposed that acquaintance be understood to mean that there is a "suitable" causal relation holding between me and Damasio, or between me and facts about Damásio. Whatever the details, the idea is that this relation holding between me and Damasio simply does not hold between me and Bush (or between me and facts about Bush). Following this line would drag us into metaphysics and the philosophy of mind, but the point here is that, depending on what view of propositions and causation is advanced, you could get to a view that knowledge by acquaintance is really a species of propositional knowledge (our third sense, below), which would trim the list of kinds to two rather than three (e.g., Sosa in [Bonjour and Sosa 2003]). Nevertheless, this is a technical point and remains non-standard. What is important for us is to recognize that, regardless of the ontological characteristics of 'knowledge by acquaintance', traditional analytical epistemology largely ignores it. And we will too.

2. that S has *an ability* to x, such as 'Greg knows how to ride a bike'. Know-how is distinct from propositional knowledge, since I can know how to ride a bike without knowing facts about bike-riding, such as that I must both peddle and turn in the direction of my fall. Know-how is distinct from knowledge by acquaintance, since I can be acquainted with a bike without knowing how to ride it. And know-how is distinct from propositional knowledge since most people—certainly children—know how to ride bikes without knowing that the must turn in the direction of their fall while peddling. Traditional analytical epistemology largely ignores this kind of knowledge as well, however the topic has received some attention lately in the philosophy of science. It arises here by way of giving an account of the epistemic merits of scientific inquiry. The idea here is that the methods of inquiry that science introduces and refines is as an important an intellectual contribution as the understanding of the natural world its theories provide for us: it isn't just what science tells us about the world that is valuable, but what it tells us about how to manipulate the world to yield facts, results, knowledge.

3. that S knows *that* x, Such as 'Greg knows that Lisbon is south of Porto'. This is traditionally called *propositional knowledge* or sometimes *knowledge-that*. Propositional knowledge is what traditional epistemology studies.

Some things to notice about propositional knowledge. Here, 'x' is a proposition, not some kind of relational structure between the agent and another object, and not an ability or skill. The basic idea behind propositional knowledge ascription is that S believes that x is true, x in fact is true, and S has a (good) reason for holding x. So, the idea is that the sentence 'Greg knows that Lisbon is south of Porto' is a case of knowledge if and only if Lisbon is south of Porto, Greg believes that Lisbon is south of Porto, and furthermore Greg has some (good) reason for holding his belief that Lisbon is south of Porto. It is debated what having a *good* reason comes to. But it isn't a total wash. If I justified my belief about Lisbon being south of Porto by consulting the pattern made by coffee grounds in my cup or rubbing the bumps on my head, we'd say something was amiss and that I don't know: lucky, crazy, but not in possession of knowledge about Lisbon's location *vis a vis* Porto. Unpacking why this is so is one of the central issues in epistemology. I view it as *the* central issue. Regardless of its rank, the issue turns out to be very subtle. One of our goal here is to get into a position to appreciate why it is so subtle.

In what remains I will use 'knowledge' to denote propositional knowledge unless otherwise noted.

1.2 The traditional analysis of knowledge:

We sketched the traditional analysis of knowledge already in our most recent comments. However, let's be a bit more specific. Traditionally, Knowledge is Justified True Belief (JTB). For any agent S and proposition p, the JTB analysis of knowledge holds that,

An agent S know that p if and only if:

1. p is true,
2. S believes that p,
3. S is justified to believe that p.

Let's unpack this analysis a bit more.

1.3 A word about truth

The notion of truth has gotten a bum rap in the twentieth century. The view from traditional epistemology is that most of the criticism directed at truth (or Truth, or Absolute Truth, or "truth", if you prefer) is misplaced, largely based upon a confusion in terms. To be sure, there are philosophically tricky problems concerning truth—just what are its semantic properties, how to handle the semantic paradoxes—which are a central topic in the philosophy of language (see [Soames 1999]). There are also tricky problems about our ability to grasp the truth, how we may be sure that we have it in our possession, or whether we ever can, in principle, have it in our possession or verify that we have it in our

possession.[1] The problems surrounding our ability to grasp the truth certainly are under the scope of epistemology: this is the classic problem of skepticism. We will return to skepticism toward the end of the lecture.

First, truth. The word 'true' is both a logical (0-place) predicate and a grammatical adjective that takes nouns and noun phrases as subjects. Grammatical examples include:

> That is true.
> It is true that Bush said that.
> The proposition that Mona is beautiful is true.

This grammatical fact leads us to think that when we assert such sentences we typically refer to something and say of it that it is true. The thing that we refer to is called the *truth bearer*. The basic idea is that to say of a truth bearer that it is true is to predicate truth of it. Here, then, the grammatical and logical predicate 'true' perform the same task: both predicate a certain property, *truth*.

Aristotle summed up truth by saying, roughly, that a man speaks the truth when he says what is so that it is so and of what is not so that it is not so, and speaks falsely when he says what is so is not so and what is not so that it is so.

There are some subtle points about the relationship between sentences and propositions, on the one hand, and about what exactly makes a thing a fact (i.e., how non-linguistic truth-makers perform the job of truth bearers). But, I take it that the trick is to forward a theory of truth that explains the underlying idea expressed by Aristotle's formulation and that an account that diverges from this (or worse, rejects it wholesale) is a failure or incoherent: to say of truth that it is not so and persuade us to start saying that truth is not so is so? (Please.)

I take it then that the main idea underlying Aristotle's proposal is uncontroversial. Nevertheless, common usage may suggest reasons to doubt that I am correct—that is, to doubt that it is true(!) that this proposal about truth is uncontroversial. For instance, we sometimes say "X is true for me" to mean either (i) that I believe X, or (ii) X is true *of* me. But, notice that 'Greg believes that Martians run the Pentagon' is true just in case Greg believes that Martians run the Pentagon and not where in fact the civilian staff of the DoD was born. Likewise, tallness is true of me if and only if I am tall. Vague predicates may be difficult to evaluate because it may be difficult to identify what proposition a sentence with a vague predicate expresses. But, again, this is a verification problem not an ontological one: 'Greg is tall' is true if and only if Greg is tall.

Perhaps some sentences don't have truth values. 'Santa Claus hires elves' appears to express a proposition that, strictly speaking, is false: there aren't any elves, nor is there a Santa Claus, so there is nothing hiring nothing. Nevertheless, the tale of Santa Claus has it that he used elves to assemble toys, and so the sentence does express a truth when the proposition is appropriately

---

[1]The distinction between possessing the truth, on the one hand, and knowing or verifying that the true proposition we believe is indeed true, on the other, is a critical distinction to keep in mind when considering the *problem of skepticism*. We won't dwell on this point here; count this note a tip for reading the skepticism literature.

anchored within this fictional story. 'Santa is a thin man' is false on both contexts. The important issue again is fixing precisely what *proposition* a sentence of a natural language expresses—fixing it for us to evaluate, that is. The truth value assigned to a proposition has nothing to do with our ability to evaluate that proposition.

Notice that quantum phenomenon don't present a distinct problem, either: the statement that,

> (Q) for any particle x, it is impossible to determine momentum and velocity for x at time t

is true if and only if it is impossible to determine momentum and velocity for any particle x at time t. We take Q to be a true description of our world; again, our judgment that Q is true is distinct from whether or not Q is true.

It has been debated whether every proposition is true or otherwise false. In considering the ontological ramifications from adopting *the law of excluded middle* Aristotle, again, asks us to imagine a sea battle tomorrow and the sentence 'The Athenians will win the battle'. Does this express a proposition and, if so, is that proposition true even before the battle takes place? This consideration takes us away from epistemology and into the philosophy of language, philosophical logic and metaphysics.[2] The crucial point here, again, is fixing what proposition is being expressed. The auxiliary verb 'will' signals that we're discussing an event in the future, after all. Nevertheless, note that our remarks are independent of this consideration. Propositions are true or they aren't true could mean that they are true or false, on the one hand, or true, false or indeterminate on the other. Call this latter view *the modest view* of truth.

Notice that even on the modest view of truth, the scope of indeterminacy is restricted to propositions whose truth makers involve events that are not determined. Furthermore, it is not clear what the structure is of a proposition expressed by sentences referring to events that haven't yet occurred. One might hold that propositions aren't intrinsically indexed by time, unlike sentences or assertions, and so determinism is avoided. 'Athens wins the battle' is true if and only if Athens wins the battle. One reason that we can't evaluate the proposition before the battle is that the truth makers aren't settled. Still, this proposition is true or if and only if Athens wins the battle.

The issue is sticky, to be sure. What is important to keep in mind for our purposes is that it is very hard to see how one can coherently maintain something *weaker* than this modest view concerning truth—which is not the same thing as saying that this view (or a stronger view) is an easy thesis to defend.[3]

---

[2]This is one key piece of the classic "free will" problem.

[3]Perhaps this is the mark of a philosophical problem: On the one hand we find complete incoherence—paradox or skepticism, say—that is nevertheless very difficult to avoid falling into when advancing an account, in this case of truth and knowledge, respectively. However, what is sometimes forgotten is that very often the *incoherence* of offering a defense of one of these pitfalls and a *bona fide* philosophical position, for one has to help-himself to some bit of coherence to advance a view. The reason serious-minded people do this is to see if they might be able to pin-point what goes wrong with the positive account: it is an intellectual exercise of failure analysis.

1.4 Another word on truth: skepticism

The following is taken from Scott Soames [Soames 1999]. There are at least five forms of skepticism about truth, which should at the least be distinguished from one another. Soames presents the distinction as follows:

1. Truth is indefinable.

2. Truth is unknowable and epistemologically dispensable.

3. Truth is irreducibly metaphysical and hence not scientifically respectable.

4. There is no such property as truth; truth predicates are trivial and lack content.

5. The notion of truth, as we ordinarily understand it, is paradoxical and thus must be either abandoned or revised.

The received view in analytical philosophy of language is that the arguments *for* these skeptical conclusions are at best inconclusive. See (Soames 1999) for a discussion of each.

Let's focus on the unknowability objection, which is a common objection raised in the context of epistemology. Again, I follow Soames.

Hans Reichenbach offers an epistemological argument for skepticism about truth in his *Experience and Prediction*, writing at the end of chapter 3 the following conclusions about truth:

> Throughout the first chapter we entertained the presupposition that propositions about concrete physical facts, which we called observation propositions, are absolutely verifiable. A more precise analysis showed that this conception is untenable, that even for such statements only a weight can be determined. With the object of obtaining more reliable statements, we then introduced impression propositions; throughout the second chapter we upheld the supposition that at least these propositions are capable of absolute verification. We have discovered now that even this is not tenable, that impression propositions also can only be judged by the category of weight. Thus there are left no propositions at all which can be absolutely verified. *The predicate of truth-value of a proposition, therefore, is a mere fictive quality; its place is in an ideal world of science only, whereas actual science cannot make use of it.* Actual science instead employs throughout the predicate of weight. We have shown, in the first place, that this predicate takes the place of the truth-value in all cases in which the latter cannot be determined; so we introduced it for propositions about the future, so long as their events are not yet realized, and for indirect propositions, which remain unverified for all time. We see now that all propositions are, strictly speaking, of the latter type; that all propositions are indirect

propositions and never exactly verifiable. *So the predicate of weight has entirely superseded the predicate of truth-value and remains our only measure for judging propositions.*

If we, nevertheless, speak of the truth-value of a proposition, this is only a schematization. We regard a high weight as equivalent to truth, and a low weight as equivalent to falsehood; the intermediate domain is called 'indeterminate'.

The move Reichbach makes is to observe that empirical propositions can never be established with absolute certainty and to conclude from this observation to the skeptical conclusion expressed in this passage.

Consider the seduction in this line of reasoning, again following Soames.

1. If the proposition that p is an empirical proposition, then what one is committed to in virtue of assertively uttering

    a. It is true that p

    that the proposition that p is true is stronger than what one is committed to in virtue of

    b. It is highly probable/confirmed/supported/verified that p.

    In each case, one is expected to have strong supporting evidence that p. But in the case of (a.), this is not enough since the speaker is also committed to the proposition that p. If it turns out that it is not the case that p, then one who uttered (a.) has made a mistake; this is not always the case with (b.)

2. Thus, the statement expressed by (a.) is stronger than the statement expressed by (b.).

3. The strongest statement one can justifiably make regarding any empirical proposition is that it is highly probable, confirmed, supported, or verified.

4. Thus, one is never justified in making the statement expressed by (a.). Since empirical truth is unattainable, the truth predicate has no legitimate place in empirical science.

Comments:

The conclusion is too broad. We know *a priori* that p iff p and the proposition that p is true iff p. Soames argues: "So, if we were never justified in asserting that the proposition that p is true, then we would never be justified in asserting p for any empirical proposition. In short, if scientific methodology excludes truth, then it excludes all empirical statements. Since this is absurd, the conclusion (4) is false.

Truth and Certainty. The main problem with the argument is that it conflates truth with certainty. Example: For any proposition p, p is probable/confirmed to degree n if and only if the proposition that p is true is probable to degree n. This is not the case with certainty. Soames, again: "The probability that the coin in my hand will come up heads the next time I flip it is, let us assume, .5. Thus the probability that it is true that the coin in my hand will come up heads the next time it is flipped is also .5. By contrast, the probability that it is certain that the coin in my hand will come up heads the next time it is flipped is clearly *not* .5. Therefore truth must be distinguished from certainty."

Assertability conditions: The first problem is in step 2. The *act* of asserting a proposition entails that one has good evidence for that proposition. However, *what one asserts*—the proposition that it is true that p—does not entail anything about one's evidence. The proposition asserted neither entails nor is entailed by the proposition that one has good evidence that p. Hence, the two propositions are distinct. Hence, (2) is false.

Soames lauches an argument against premise 3 that focuses on the semantics of *strength* with respect to probability or confirmation measures assigned to a proposition. (See Soames, pp. 32-3). The idea is that of the options we have for unpacking this notion either involve an equivocation in meaning, between one notion used in premises 1 and 2, or a consistent analysis that renders the premise false.

This discussion is far from a knock-down refutation of the epistemological-based skepticism to truth that attended the logical positivist tradition. But one should know that *verificationism* (the view of truth that Reichenbach alludes to in this extended quotation) is widely discredited. There is considerable work for a truth predicate in linguistics, not to mention logic, to which verificationism has thought to be ill-suited. And once one concedes these points—and thereby discredits a *general* theory of verificationist truth—it becomes difficult to defend the view piecemeal against modest (but general) proposals such as the one advanced here in the lecture.

1.5 A note on belief

There is work at the border of epistemology and the philosophy of mind on the structure of belief, how we recall one, and what state we must be in to have knowledge—actually, the action here mostly concerns what state we must be in for a belief to be justified.

As a first pass, notice that there are a few different kinds of beliefs and a few different ways to form a belief. First, there is a distinction between fully conscious beliefs and those that are in one's subconscious. For instance, there seem to be things we know but where the belief is subconscious rather

than conscious. You believe that Zebras have stripes and live in Africa. That is, you believed it before I mentioned it. It wasn't the case that I told you something new, rather, you simply recalled a belief you already have. What you are doing now (if still thinking about Zebras) is entertaining conscious beliefs about Zebras.

Second, there is a difference between beliefs that an agent can explicitly formulate and beliefs he cannot. You might have beliefs about the smell of a loved one's hair or believe that you can identify a cello when you hear one. Whether such beliefs are conscious or subsconscious, notice that you needn't be able to explicitly formulate the content of either belief—that a lover's hair smells like such-and-so, that a cello sounds like this rather than that—to have a belief.[4]

Must you be able to recall something to consciousness that you believe in order to know it? Epistemologists tend to answer No. What, then, are the standards to apply to determine when an unconscious belief is nevertheless known? What are the conditions necessary for a belief to serve as justification for another belief that is known? What is the nature of these relations? And so on.

One point to bear in mind is that answers to these questions are driven by what theory of justification one adopts.

---

[4]An aside: There is some work at the borders of the philosophy of mind and cognitive psychology that propose doing away with the notion of belief altogether. Paul and Patricia Churchland [Churchland 1981] are the most notable proponents of this eliminativist view concerning belief. They draw from neuroscientific evidence to make their argument to the effect that 'belief' denotes a concept that is part of folk psychology, and so should be replaced by respectable notions grounded in contemporary cognitive psychology and neuroscience. Taking this issue up is well beyond the scope of this course, however I wish to make the following comment. There is little doubt that we have a pre-theoretic notion of belief: we hold doxastic states and make judgments about the contents of those states—whether, for instance, the proposition is true, is well justified, is moral, is beneficial, is a consequence of another proposition, and so on. It may be that talking about this activity in these terms—about a state of "mind" having "content" that is a "proposition" such that this mental object may be evaluated along these parameters—is wrong headed, despite the intuitive pre-theoretic appeal. And, so, perhaps we should turn to another framework to explain this phenomenon. This is a plausible position to advance. The point to notice about the Churchlands' view (*eliminativist materialism*), however, is that they are advancing a far more radical thesis—that since neuroscience doesn't posit the existence of unified 'belief states' then we should conclude that belief doesn't exist and pitch our pre-theoretic notion of belief in the same bin as notions like fairies, witches and goblins. We should hold out for a more sophisticated theoretical account of our mental life, advise the Churchlands, and not try to construct accounts upon "folk psychology".

The trouble is that we evaluate states, from our first person perspective. Evidence taken from neurological accounts of brain activity then is only of indirect relevance. It won't do to give up on providing a suitable account of what it is—our doxastic states—that we make judgments about from our first-person perspective based on the third-person study of the brain since this evidence, by its nature, has no room for such first-person states. We're not all going to stop having beliefs—or whatever you'd like to call them—even if we all decided to believe(!) that the Churchlands are right after all. In sum, even if folk psychology is wrong, it doesn't follow that mental phenomenon (including first person doxastic states) don't exist.

1.6 A note on evaluating beliefs

Beliefs can be evaluated on other dimensions than on whether the content of the belief, the proposition, is true. Someone might have a huge financial stake in a biotechnology firm and thereby believe that the drugs under development are a success; I may benefit in recovering from a medical operation more rapidly if I believe in God than if I don't; A mother's love for her son may make it impossible for her to believe that he committed a murder he indeed committed, to which he confessed, and for which he was convicted in a court of law. And so on.

This is to say that various evaluative notions may be applied to beliefs. The central evaluative notion that concerns epistemologists is *justification*. The shareholder's avarice has no (direct) bearing on the bio-chemical effectiveness of drugs; a mother's love has no (direct) bearing on the facts of her son's case; finally, while there is a correlation between a patient's belief that he will get better or be well taken care of and his actually recovering, the practical benefit associated with this belief should not be confused with the epistemic merit of the belief. (On this last case, see pragmatic justification below).

To begin to pin down what we mean by 'epistemic merit', consider for comparison a few distinct senses of justification.

1. *Moral justification* arises when a belief is morally justified. Consider the belief that torturing newborn babies for entertainment is wrong. The belief is morally justified, setting aside for the moment how this evaluation is established. (This would take us into ethics, after all.) That is, it is morally justifiable to judge the torture of newborn babies as wrong. Contrast this with reading this sentence to express (albeit awkwardly) an epistemic evaluation. The proposition for consideration here is whether there is no case of someone torturing newborn infants for entertainment. If there is reason to think this is true, then the claim that it is unjustified to believe otherwise is justified.

2. *Epistemic Justification* applies to beliefs and, perhaps, other cognitive attitudes. It is a necessary constituent of knowledge, having something to do with truth. Unpacking this idea accounts for the bulk of the work that is carried out in epistemology. It may be helpful, however, to contrast this notion with still another notion of justification, *practical justification*.

3. *Practical justification* arises when a belief has instrumental value, such as the patient's belief in God helping along his recovery. A clearer example of pragmatic justification, perhaps because it is less controversial, is the case of a baseball batter's belief that he will get a hit when he's up at the plate. We know that the very best baseball players in the history of the game get a hit less than 1:3 times at the plate; hence, it is not rational for him to believe at each bat that he will hit the ball. That is to say, such a belief is without epistemic merit. However, the batter is pragmatically justified in believing that he will hit the ball—indeed, his

believing so likely contributes to improving his performance.[5] Still, this doesn't change that he'll be called out two-thirds of the time at the plate.

In contrasting *epistemic* justification with *pragmatic* justification, the key idea is that benefiting from believing something has nothing to do with whether I know it is true, and so has nothing to do with whether I am epistemically justified in believing it.

1.7   A brief introduction to justification

The bulk of work in epistemology concerns theories of epistemic justification. (I will omit mention of 'epistemic' from here out). Indeed, the bulk of this course looks at the main theories of justification to have been proposed and defended. There are a couple of principles to keep in mind, which we will label for future reference.

1.7.1   Two Principles of Justification

A. A belief can be justified and false. (*fallibilism*)

B. Demonstrating that a belief is justified is not a necessary condition for that belief to be justified.[6] (*non-verificationism*)

1.7.2   Sources of Justification

An agent S at a particular time t may have justification for p drawn from one of three sources, namely

1. S's beliefs

2. S's sensory experiences

3. The world external to the subject's psychology

Thus we have an sketch of the main conceptual components of knowledge.

Historical aside: An important point to raise here is that many of the more radical 20th century epistemological ideas to leak into wider intellectual circles—constructivism with respect to truth ("truth"), relativism with respect to epistemic standards (feminist epistemology, for instance) or post-modern pronouncements of the end of epistemology (e.g., Richard Rorty)—are reactions to the failures of mid-century analytical epistemology, particularly logical positivism (e.g., Russell, Ayer, Carnap). Ironically, these critiques now form the main

---

[5]But his optimism does not change the epistemic status of the belief. What we've articulated here is the limits of wishful thinking.

[6]Example: Suppose S has a headache. Having a headache is good evidence for S believing that he has a headache. Failing to convince S's doctor, boss, wife, or anyone else that he has a headache does not entail that he is no longer justified to believe he has a headache. For, he has a headache.

menu from which mainstream *intelligencia* choose. Call this "The New Cynicism". The trouble is, these views don't stand up very well on their own.[7] The very idea of sentences expressing propositions is jeopardized by constructivism, a point we touched upon when discussing truth. At the very least, it takes a considerable degree of subtlety to work out a coherent constructivist view—the details of which you simply do not generally see present in the new cynicism. (Indeed, the best model for constructivism is still Kant, who's metaphysics is nevertheless far more constrained than than pop constructivism.) Another constructivist, Thomas Kuhn, spent most of the remainder of his career distancing himself from the two ideas that are now associated with his name: *paradigm shifts* and the view that reality is a social construction. Alas, Kuhn was no Kuhnian.

The point? Resist decadence. Intellectual cynicism is, thankfully, as flimsy a view now as it was in the Hellenistic days. There is an intellectual dishonesty to at once corrode the foundations of inquiry—evidence, sound argument, rationality, truth—while relying on these foundations to advance an argument to explain their poverty and persuade us to give up. It is philosophy on the cheap: failing to give an explanation for p does not entail that you've an explanation for $\neg p$.

2 Lecture Notes: Gettier Counter examples

One of only two papers ever published by Edmund Gettier, a two page article "Is knowledge justified true belief?" [1963] offered an incisive counter-example to the JTB analysis of knowledge. Epistemologists have been exercised by Gettier-style counter examples since. One of Gettier's two counter-examples runs as follows.

> Suppose Smith has very strong evidence for the proposition A, *Jones owns a Ford*. Smith's evidence might include that Jones has always owned a car in the past, it has always been a Ford, and that Smith has just accepted an offer of a ride from Jones who is driving a Ford. We are then asked to imagine another friend of Smith, Brown, whose whereabouts are completely unknown to Smith. Smith selects a place at random and entertains the following proposition B, *Jones owns a Ford or Brown is in Barcelona*. Since A entails B and let us suppose that Smith grasps this entailment, Smith is justified to believe B. But now imagine that in fact Jones does *not* own a Ford; the present car he is driving is a rented car. Furthermore, by chance, suppose Brown is in Barcelona. So, B is true. However, it no longer appears that Smith knows B.

---

[7]Indeed, epistemologists seem content with the published record discussing these views during the 60's and 70's. The field now largely ignores these issues, despite their prominence in broader culture in second and third source literature. An exception is Thomas Kuhn, who is still discussed seriously. But these discussions are confined to the philosophy of science, not epistemology.

2.1 Replies to Gettier

Summary of options:[8] (i.) One may focus on the notion of epistemic justification while rejecting one or both of Gettier's assumptions about such justification, thereby undercutting the counter-examples. One may in turn (ii.) accept Gettier's reasoning and try to minimize the damage by downgrading the importance or scope of the counter-examples, either by offering a substitute analysis or by tossing out the project of analysis altogether.

1. Rationalists deny fallibilism [principle A, above] that there can be a false justified belief. This move blocks (standard) Gettier cases by denying that false beliefs can be justified. The problem is that this results in notions that are too strong—'super-justification' and, in turn, 'super-knowledge'.

2. Reject the JTB analysis by pitching justification altogether. Crispin Sartwell has recently argued [Sartwell 1992] this view. In reply to the objection that lucky guesses don't count for knowledge, Sartwell replies that lucky guess don't amount to true belief: his idea is that no belief is held in isolation but must be surrounded by a whole set of beliefs. It isnt so simple to chance upon collections of beliefs by lucky guesses. His positive argument for TB over JTB: Our ultimate aim in inquiry is knowledge. Either we want justification in our beliefs because it is a means to truth or justification with truth as the aim of knowledge. But if justification is of only instrumental value, then it is not part of the aim of inquiry, and so not an essential feature of the concept *knowledge*.

   We might question whether from observing that the aim of inquiry is true beliefs we should infer that knowledge is true beliefs. The aim of scoring in football is to put the ball in the goal. However, we (ok, not me: you) are intimately interested in the skill of players who shoot on goal. To borrow an example from Sosa, we readily judge the difference between a winning shot and a lucky shot—the former is a result of skill and cunning, say, while the other a matter of chance—and are interested in the development of this skill, assessing a player's ability as well as his point total. It is part of the design of the game that there is a high correlation between skill and goal totals. Returning to knowledge, we intuitively take there to be some correlation between sound epistemic practices (paying attention to available evidence, withholding judgment when without sufficient evidence, updating our beliefs on the basis of what our evidence suggests) because we think that these practices are correlated with getting true beliefs. Just as the concept of sport includes a clear success component and a built in preference for skills that maximize success, so knowledge has a success component (true belief) and also an interest in the skills that maximize success. A principle difference is that sports include rules to favor some practices over others (kicking the ball with spin, say, over murdering

---

[8]For a more detailed treatment of the possible replies to the Gettier problem, see [Shope 1983].

your opponent's goalkeep) whereas we haven't the luxury to step out of ourselves for a god's eye view. Pity, that.

3. *Justificationist strategies*: Strengthen the justification condition. The idea behind this strategy is to strengthen the justification condition by requiring, for instance, that an underlying inference or argument not depend upon a false lemma or by requiring that the justification not be 'defeasible' by any falsehood. We turn next to theories of justification that attempt to unpack this notion of justification.

3 Introduction to Theories of Justification

The main work in contemporary epistemology is on theories of justification. By characterizing what it means for a belief to be justified (*rationally accepted*), epistemologists aim to construct accounts that explain the constructive role that justification plays in securing knowledge—its truth conduciveness; its role in evaluating our own and other people's beliefs; how it is passed from belief to belief by reason or memory, or passed from the world to belief through our direct experience of the world or talking with others; and how in principle to distinguish between how to do all of this while avoiding the ancient problem of skepticism and the more recent Gettier-style counter-examples.

The notions that we've laid out above begin to be difficult to keep clearly in mind as we move forward from here, which is perhaps another feature that philosophy shares with mathematics: both, at bottom, are formal disciplines.

This said, there are a couple of ways to proceed from here. If you crack open a textbook on epistemology from 7 years ago or longer, you will likely see a claim that there are two basic approaches to theories of justification: *coherentism* and *foundationalism*. Let's sketch these before turning to how the subject is viewed today.

> **Coherentism**. A belief is epistemically justified in virtue of being a part of a coherent body of beliefs that are appropriately comprehensive.
>
> The intuitive picture to associate with coherentism is that of a raft at sea. Just as a raft succeeds by having planks that, together, provide buoyancy to keep you from drowning, so too does a consistent body of belief, taken together, provide justification for thinking that that body of belief is (more or less) true.
>
> 1. The technical problem with coherentism is spelling out what *coherence* and *comprehensiveness* each amount to. Is it sufficient for the contents of beliefs to be logically and probabilistically consistent? Doubtful: We may take an arbitrary set of sentences, close them under consequence, yielding a logically consistent set. Likewise with an initially consistent assignment of probability to an arbitrary set of sentences.

If comprehensiveness isn't a matter of the cardinality of a set of consistent beliefs, then what must we add? Well, we might ask that a set of beliefs not only cohere with one another but also cohere with one's own relevant experiences. But this suggestion is at odds with coherentism itself, which maintains that coherence and comprehensiveness are *sufficient* for epistemic justification.

2. The fundamental problem with coherentism is that there is no principled explanation offered for how coherence explains the difference between true belief and knowledge. Paranoids have very coherent stories to tell about conspiracies that aren't true—this is what it means to be a paranoid, after all. Keeping one's story consistent seems too weak a demand for a belief to be justified. What is missing is the requirement that one's coherent set of beliefs is suitably connected to one's experience.

3. Let's make sure we are still together on the concepts in play. It is important to distinguish coherentist views of *justification* from conherentist views of *knowledge* and, finally, coherentist views of *truth*. The list is ordered by decreasing degree plausibility (or increasing degree of confusion, if you like). The most promising coherentist view of knowledge would have it that truth has no role in the proper analysis of knowledge, namely that knowledge is just justified belief (JB). The problem here is that there isn't a good reason to collapse the distinction between *rational acceptance* and *knowledge*; we have a clear idea what we mean by each, and holding the two synonymous is an abuse of language. A coherentist view of truth would hold it that a consistent set of beliefs is true. But, in the case of empirical beliefs, this leads to an incoherent ontological relativism—one where there is no difference, in *principle*, between a room full of paranoids whose beliefs are each self-consistent but pairwise inconsistent. In general, it is a hard sell for epistemology to drive metaphysics.

Foundationalism. A belief is epistemically justified in virtue of being a direct recording of what is present in the mind—by experience or, more generally, a state of consciousness—or else is based upon such foundations.

The intuitive picture to associate with foundationalism is that there is some class of foundational "blocks" upon which a superstructure of beliefs rests: in this case, the planks of wood are arranged to form a house standing on solid ground.

The first thing to record about foundationalism is that there are several varieties. Philosophers disagree upon what should count as foundational elements, how the class of foundational elements get

their justification status, what the suitable relations are between the foundational elements and the structure of beliefs resting upon those elements, how justification is propagated from foundational elements to other points in the structure, how it is propagated within the structure, whether that propagation is monotonic, non-monotonic, and so on. Indeed, we could spend a semester alone talking about foundationalist theories of epistemic justification. But let's follow our model here of giving a brief introduction.

1. Given the varieties of foundationalism, it is a bit more difficult to get a handle on *the* technical problem for foundationalism. Let's set this discussion aside, then, until we take up each theory in more detail.

2. *Rationalist Foundationalism*: The idea here is that only rational intuition can provide a secure foundation for knowledge. Axiomatic systems serve as a model here. Descartes *Cogito* argument is another: "I am a thing that thinks, so I am a thing that exists". The rational insight Descartes had that he was a thinking thing provided, he thought, sufficient foundations from which to account for how he possessed knowledge.

   The failures of rationalism are evident in Descartes *Meditations* and, more recently, Frege's logicist project (the reduction of all of mathematics, including to set theory, to logic).[9] The main problem is that the scope of our knowledge appears to be broader than what is permitted by the austere foundations allowed by a rationalist program. In the case of Descartes, we have empirical knowledge claims with which to contend. In the case of logicism, we have the limits of expressibility of first order logic (FOL) and the unwieldiness of higher order logic (HOL). Another problem, restricted to Cartesian foundationalism, is that the view denies fallibilism (principle [A] from above), by demanding that a belief is justified only if we're not mistaken. But belief that p is true itself is *true* only if we're not mistaken that p, and so we know p only if we're not mistaken. Nevertheless, we may have very good reason to believe p is true and so be justified to believe that p.

3. *Empiricist Foundationalism*. Rationalist (Cartesian) foundationalism runs into trouble by holding (i.) we are infallible about our own internal belief states, (ii.) that every justified belief state depends on beliefs about internal belief states, and (iii.) that deducton is the only way to confer justification from foundational beliefs to non-foundational beliefs. *Empiricist fondationalism* may be thought of as an expansion of Cartesian foundationalism, one that addresses these three short falls.

---

[9]This isn't dead history: current work on second order logic among philosophical logicians is part forms the basis of neo-logicism.

In addition to rational intuition, present sensory experience or memories of past sensory experiences are included as bearers of justification. (Indeed, the classical sources of "the given" are *intuition*, *introspection*, and *observation*.) Empricism also tends to be more liberal with respect to the inference steps it allows: in addition to deductive reasoning, empiricism typically accepts inductive reasoning as well.

The problem with empiricist foundationalism again is tied to the observation that the scope of our knowledge appears to be broader than even these foundations provide for. Quite a bit of our knowledge is not obviously tied to either rational intuition or tied to immediate or recalled sensory experience. Omitted from explanation is one's knowledge of history and much of science, since it is doubtful that we each have adequate sensory experience to tie our knowledge to the relevant empirical events sufficient for justification (on this view). Another problem is the distinction between the class of foundational beliefs and non-foundational beliefs. What are the conditions for a belief to be self-justifying? (Or, more subtly, what are the conditions for a belief not to require justification from another belief. This distinction arises from reflecting upon basic experiences justifying beliefs. Having a headache is good grounds for the belief that I have a headache, and so justifies it. The headache isn't a belief, nor is it an experience that needs any other (epistemic) justification.

4 Internalist and Externalist Theories of Justification

We remarked in the last section that there is more than one way to classify theories of justification. The first, traditional way, is to mark the distinction between *coherentism* and *foundationalism* in much the same manner that we just sketched.

However, recent attention has focused attention not so much on the coherentism vs. foundationalism debate than on something called the *internalism* vs. *externalism* debate. We are now in a position to see the structure of this dispute.

Recall again the doctrine of *The Given*. Classical, empiricist foundationalism holds that a belief may receive justification in one of three ways, namely by

1. *rational intuition*,

2. *introspection*, and

3. *observation*.

It is not clear how to unpack these notions today because there are disputes in the philosophy of mind and in epistemology that make classification schemes unstable. One reason for the instability is that there is disagreement about concept priority—whether features of language should take precedence over features of mental states, whether particular empirical facts about psychology should take precedence over traditional evaluative notions, and so on. The approach here will be to describe the main disputes in epistemology that are contributing to this instability. The aim is to give a bit more detail to the fundamental issues involved in proposing an adequate theory of justification.

The debate between internalist theories of justification and externalist theories of justification may be characterized as a disagreement over the role that these three sources play in having a justified belief. The (very) rough picture to call to mind is that internalist theories of justification are predicated on the view that the first two constituents (intuition, introspection) are sufficient for an adequate theory of justification, while externalist deny this and hold that important mind-independent facts about the third (observation) are necessary.

Hence, one way to understand these categories is to think of internalist theories as defending a historic tradition (at least since Descartes) of restricting the menu of sources of justification to things purely in the agent's head—the mental states corresponding to his rational intuition and his memory—while externalists think that this is insufficient for justification, particularly for accounting for how empirical beliefs receive justification.

'Externalism' denotes a relatively recent class of theories—the first externalist theory, *reliabilism*, was proposed in the mid 1960's by Alvin Goldman. There is considerable disagreement among externalists about what is missing from the classical, internalist approach. But the property that externalists theories share is the inclusion of an extra-mental condition for justification, a condition that contributes to the justificatory status of an agent's beliefs but that is impossible, in principle, for the agent to grasp by reflection. In short, we have this:

> The *internalist* views the justification status of a belief to be determined (in principle) by features that are completely within the cognitive grasp of the agent who is having the belief. On this view, an epistemic agent should be able (in principle) to figure out the justification status of his beliefs by reflection. This view of justification seems better suited to satisfying the *evaluative* role that justification plays; it traditionally has trouble accounting for the *truth-conduciveness* role that justification is thought to play.

> The *externalist* views the justification status of a belief to be determined (in principle) by features that are not completely within the cognitive grasp of the agent having the beliefs. A necessary feature of a belief being justified, on this view, is that the belief be formed in an "epistemically virtuous" manner, one that is likely to yield true beliefs. There are various proposals for specifying epistemically virtuous belief formation. The important point to keep in mind is that

the function of this belief forming processes is not a property that can be directly grasped by introspection. This view of justification seems better suited to satisfying the truth-conduciveness function that justification plays; it has trouble accounting for the evaluative role that justification plays, since it includes a condition for being justified that is in principle not accessible by internal reflection.

The *problem of skepticism* enters into the discussion here.

4.1   A word on skepticism

The problem of skepticism is a mind bender. The idea, at its core, is to ask us to *imagine* that we are in a world unlike this one, the actual world that you and I experience. In this other world we're provided the same mental life that we have now—by a demon messing with our senses in that world, or by us being a brain in a vat (BIV) and electrochemically deceived in that world. In other words, all the sensations, thoughts, feelings—everything making up our mental life in that world is as it is in this world. But in that BIV or demon world there is no actual world "out there". The question is, how do we distinguish between that world and the one that we're in. The problem of skepticism, then, is a hard diagnostic tool to test a theory of justification.

There are two varieties of skepticism that are important to keep distinct.

1. *Pyrrhonian Skepticism*: Pyrrho of Elis (c. 365-270 B.C.) held that the reasons in favor of a belief are never better than those against it and so we shouldn't have beliefs. His advice was to stop worrying and just accept the apperances of things.

   But consider the principle:

   > (Pyr) the reasons in favor of a belief are never better than those against it.

   But what about the belief that (Pyr)? Are we to regard the belief *that the reasons in favor of a belief are never better than those against it* a belief that itself has no better reasons in favor of it than against it? What are we to make then of Pyrrhonian ethics, the prescription to not worry and "accept" the appearances of things? Presumably there are at least two beliefs that we're presumed to have: that (Pyr) is true and that the correct prescription for action in light of (Pyr) is to not have beliefs but to accept the appearance of things. What should be believe either of these?

   We know about Pyrrohian skepticism from Sextus Empiricus (A.D. 200). The lesson of Pyrrhonian skepticism that philosophers take is that skepticism must help itself to a reasonable belief in *something* on pain of incoherence: (Pyr) is false. But this point wasn't clear until the 17th century, when René Descartes wrote his *Meditations*.

2. *Cartesian Skepticism*: Cartesian skepticism is more modest, for it concerns the possibility of having *knowledge*, not the possibility of justified belief. It is important to note that the Cartesian skeptic assumes that the world is as it appears to be, but then raises the possibility that it is radically different and then asks on what grounds could we sort out the perverse case from the actual case. The important point is that Cartesian Skepticism posits that we don't know that we are decieved; it does not posit that we have no reason to think that we aren't deceived.

   The twentieth century BIV version blocks some easy solutions (including Descartes own cogito proposal) by putting "us" in a situation were it is just our mental life that stays the same in both situations and not, perhaps, extra-mental characteristics that accompany experience. (See? Who says there is no progress in philosophy.)

The serious question the problem of skepticism raises is how to construct a theory of knowledge that is strict enough to distinguish between genuine knowledge and wholesale deception but weak enough to include empirical knowledge. One reason that BIV-like thought experiments are useful is that they are engineered to filter out irrelevant indices that contribute to 'justification' getting various semantic assignments but that distract us from isolating just those properties that produce genuine disagreement over their inclusion to the concept of justification.

Consider: we may concede that there is something correct or rational about the BIV's beliefs: from his point of view, he does the best that can be expected of him. On the other hand, adopting this line runs the risk of completely abandoning the truth-conduciveness role of justification; we may then reasonably ask what good *is* doing one's best if it has no bearing on the truth of matters. It is here, if doing philosophy, that one takes a closer look at the constituent notions to see if a finer analysis might yield a collection of properties that accounts for these two desiderata and provides an adequate explanation (according to intuition) to various cases presented to the theory.[10]

4.2   Epistemic Statuses

Let's try to provide a little more detail about the relevant senses of justification in dispute. A recent proposal of Ernest Sosa will serve our interests. Sosa asks us to distinguish between three epistemic statuses, called *V-APT*, *V-ADROIT*, and *FA*.

First, a virtue epistemology V includes two features:

Va: X is an intellectual virtue only if X would produce a high ratio of true beliefs,

---

[10]Another question that comes into play is principle [A], *fallibilism*. Reject [A] and you pitch empirical knowledge; keep [A], and you need to take a harder look at what that principle comes to as well.

Vb: B is a justified belief only if B is a belief acquired through the exercise of one or more intellectual virtues.

Another way to interpret Vb, where w ranges over possible worlds, is *J-APT*: For all w, B is apt-justified in w only if B is acquired in w through the exercise of one or more intellectual virtues that are virtuous in w. There is a problem with J-APT that we won't dwell on here, but it leads Sosa to propose another principle, adroit justification *J-ADROIT*: For all w, B is adroit-justified in w only if B is acquired in w through the exercise of one or more intellectual virtues that are virtuous in our *actual* world.

We are now in a position to view the distinction Sosa wishes to draw our attention to. If *Va* is added to *J-APT* and J-ADROIT, we get the following virtue epistemic properties.

V-APT: For all w, B is *apt-justified* in w only if B derives in w from the exercise of one or more intellectual virtues that *in that world w* virtuously would produce a high ratio of true beliefs.

V-ADROIT: For all w, B is adroit-justified only if B derives in w through the exercise of one or more intellectual virtues that *in our actual world* virtuously would produce a high ratio of true beliefs.

Now, what Sosa asks us to do is compare these *externalist* virtues with an internalist virtue, Foley rationality, which is roughly this:

FA: The status assigned to a belief when that belief would survive deep reflection under the agent's strongest epistemic standards.

Following Sosa, notice that someone may be Foley rational (FA) while being deeply superstitious—and so, not entirely rational.

One way that internalists reply is by posing the following question: *Do we have any good reasons for thinking that our beliefs about the world are true (or at least approximately true)?* If so, what form do these reasons take? (This, notice, is to refocus the debate by concentrating on the importance of the evaluative component of justification: have we good reasons, from our perspective, for thinking a belief to be true?)

The basic idea behind internalism is that an agent S is justified to believe p if and only if S has good reasons for p. The complaint that internalism has against externalism is that the additional features that externalism proposes cannot be recognized "from the inside" based upon the subjective character of the experiences.

Externalists think that the idea of a genuinely internalist position—one that forwards the very idea of an agent having a good reason for thinking that a belief is true that can be recognized by the agent—is fundamentally incoherent or impossible.

Internalism is difficult to defend. The failures of Descartes (rationalism) and later Locke (Empiricism) led to British Idealism (Berkley) and skepticism

(Hume) and then to the singular achievement of transcendentalism (Kant), a return to internalist foundationalism phenomenalism (Russell) and coherentism (Blanchard).

What is new about externalism is that, instead of relying on God (Like Berkeley's Idealism) or transcendental categories (Kant's solution), externalism is an attempt to get a foothold on items outside the mind without a lot of metaphysical overhead (God, the categories). One thing that is interesting about the proposal is that it allows us to focus on the very different notions of justification that emerge, since the notions that externalism appeals to (Causality, probabilistically reliable belief processes, cognitive psychology) are not metaphysically suspect.

5 A Word on Philosophical Method

No doubt a frustrating feature of contemporary philosophy to laymen is the seemingly insular nature of the field. Opening the *Journal of Philosophy* or the journal *Mind* to read convoluted arguments involving brains in vats or one's *dopplegänger* on twin earth having the same mental life as you—one can be forgiven for thinking the whole exercise not serious or a waste of time. However, thought experiments such as these *can* be put to serious use. At its best, philosophy uses these thought experiments as devices to draw insightful conceptual distinctions among notions that otherwise lead us to confusion, paradox or skepticism.

Philosophy, like other intellectual endeavors, is a human activity, by which I mean to say that it suffers the same foibles and limitations as other human endeavors. Philosophy is fallible, pursued by people who or more or less skilled, more or less honest, and who exhibit in varying degrees the virtues or properties of the areas they study: it is quite easy to find dogmatic epistemologists, unethical ethicists, and irrational logicians. 'Philosophia', the root of 'philosophy', denotes lover of wisdom not a wise person.

Another thing that might appear maddening is that philosophical views are advanced by proponents who *advocate* positions much as attorneys spar in court on behalf of their clients. Hence, one way for philosophers to gain a positive professional reputation is by ably defending otherwise crazy views, just as one way for defense attorneys to gain a good professional reputation is by ably defending clients who aren't the most admirable citizens. But whereas the tangle of conflicting interests and human fallibility make an adversarial justice system the least worst (and so, practically speaking, perhaps the best), there is reason to question whether this ancient practice of philosophical disputation serves our best intellectual interests.

The reason to doubt that this is the best way to do philosophy is that one of the virtues that the study of science and philosophy is supposed to impart is a respect for the impartial weighing of evidence and an aim for the truth. This virtue, like virtues in general, are best imparted by example rather than as the conclusion of a clever argument. As I remarked in this essay, there may

be cases in everyday life in which our actions should *not* be directly guided by epistemically justified beliefs: sometimes acting in accordance with our knowledge may be to commit an immoral act, and so we should refrain. Likewise, it may at times be in our interest to act in accordance with beliefs whose outcome will benefit us, practically, to advance our interest in a professional or personal setting—for accolades, for love, for survival.

Still, this observation does not entail that scheming and persuading is all that we do—which I take to be a tenet of what I've called the new cynicism. (Nor is the only other option to be moral scolds, standing on the sidelines of life shaking our finger). We also are interested in knowing the truth—even if, perhaps, only to devise a strategy to secure the job, the love, the glory we want. The idea here is that pressing one's advantage is predicated on having some clear idea of what one's interest are and how to manipulate the conditions at hand to advance those interests. The morality of doing so is a entirely different matter. (See the distinction between moral justification and epistemic justification). The point is that you needn't embrace piety to reject cynicism: having some knowledge of a situation is a necessary precondition to perform an intentional act, whether that act is moral, immoral, solely for our own interest, or for some cause "outside" of oneself.

Now, we may debate the extent to which people are interested in the truth, or are themselves truthful and reasonable. We may debate the extent to which particular classes of people are epistemically virtuous: the extent to which politicians are interested in the truth to make sound policy, the extent to which journalists are interested in representing the truth to their readers, the extent to which scientists are committed to epistemic virtues such as respect for evidence, an aim for the truth, and a willingness to let evidence guide belief. But no matter how dim a view we may take of the professionals engaged in one or all of these professions, it does not follow that epistemic virtues don't exist, that truth is a matter of opinion, that it is impossible to get at the truth, and so all beliefs are of equal merit. So, as a corollary, it does not following that there is no such thing as sound policy addressing the true dimensions of a conflict (however hard it may be to identify those attributes and their proper weight); nor does it follow that there is no such thing as a fair and accurate presentation of an event, either in the case of the journalist writing for a newspaper or a bench scientist writing in his lab notebook. It is easier to catch mistakes and deception in the sciences because of the common (but not universal) demand for repeatable experiments, but it does not follow from this that science is "more true" than journalism. I have suggested that talking this way leads to more confusion than its worth. Instead, the case seems to be this:

1. Beliefs can be true or false and justified. (*Rationally accepted*).

2. Beliefs can be true or false and internally-justified. (*Foley-rational*).

3. Beliefs can be true and unjustified. (Lucky guesses).

4. Beliefs can be false and unjustified. (Unlucky guesses).

The idea, in sum, is that justification can come in degrees but truth does not come in degrees. Thus our beliefs can be more or less justified but not more or less true. A belief can be close to a true belief in the sense that the content of a belief can express a proposition that is false but somehow "similar" to a true proposition. The problem is that it is very difficult to pin down this notion of "similarity", and so very difficult to make the idea of "x is more true than y" anywhere near as precise as the notions we've been working with. The relation of similarity in force here appears to be too context dependent to do work in an explanation.

Finally, a word about the results of philosophy. At its best, philosophy can show us that certain positions are hopeless, meaning that they lead to paradox, skepticism, or worse. Arguments are waged trying to show this for various positions, and some are stunning successes: Russell's paradoxes was a clear blow to niave set theory. And many more, while not decisive, certainly give us pause. Progress is slow, and on some fundamental issues—such as how our mental life is related to the world beyond ourselves—may remain unresolved. But even these difficult problems provide us with insight, if only the ability to better articulate what we don't understand.

5.1   A word about science and epistemology

One will hear (or read in referee reports, say) the use of the term 'scientific' or 'science' as a normative term of praise, such as "This paper is not scientific", meaning that it there is something blameworthy or defective in the paper. There is an understandable reason for this usage to have appeared: science ranks as one of humankind's most astounding intellectual achievements, if not the most astounding intellectual achievement. In very many fields of science, the *epistemic* authority of scientific claims serves as a gold standard by which other claims are measured.

Nevertheless, I think we should resist using 'scientific' as a term of epistemic praise. The reason is that this use is at once too inclusive and also too narrow. It is too inclusive because we regularly distinguish good science from lousy science, good experiments from badly done experiments, good scientists from bad scientists, where the evaluation in each example is epistemic. Using the term 'scientific' as a term for epistemic praise is too narrow since it exludes other perfectly acceptable modes of inquiry that exercise the critical epistemic virtues—an aim at the truth, an impartial weighing of evidence, a willingness to adjust one's beliefs in light of one's total evidence, and the like—but that clearly aren't sciences. History gives us knowledge of our past, and cooks and detectives use a method of inquiry not unlike those used in experimental sciences. However, historians chefs and police detectives are not scientists.

The demarcation of science from non-science is an exercise that attempts to define what the difference is between astronomy and astrology; between evolutionary biology and creation "science"; between science and religion. What I am suggesting here is that the constituents from which to mark out the distinction are not found at the level of the sciences—they are too varied in particular

methods suited to their problem or subject domain. Some sciences make heavy use of statistics (e.g., the social sciences, medical sciences, chemistry, biology) others hardly at all (e.g., geology, archeology); some make use of controlled experiments (e.g., physics, chemistry, pyschology) others it is impossible to run controlled experiments (e.g., astronomy, seismology). So what is it?

Karl Popper deserves credit for getting part of the answer in his claim that scientific claims are *falsifiable*. (The parts he gets wrong is that he rejected the idea that hypotheses are confirmed or accepted outright and deduction was the only rational mode of reasoning). The idea behind falsificationism that survives is that our beliefs (not just our scientific beliefs) should be open to revision. We should have some idea both *why* we hold a particular view and be receptive to evidence that would undermine the grounds we have for holding that belief. But this simply falls out from what we've advanced as epistemic virtues: Respect for the truth, respect for evidence, and a willingness to change one's beliefs to accommodate one's evidence.

No doubt some of our beliefs are more entrenched than others. Also, we each have access to different kinds, amounts, and quality of evidence for our beliefs. Consider:

> (p): Power corrupts.

A young, disenfranchised man may belief (p) and have ample reasons, stemming from his disenfranchisement, to cite his case. Likewise, another person may know of a person who came to power (elected to the U.S. Congress, let us say) and conducted his career with notable honesty and even-handedness; a counter-example to (p). We shouldn't be surprised if these two individuals fail to find agreement—even after they pin down what proposition (p) expresses and, let's suppose, that both are equally epistemically virtuous (whatever in the end that amounts to). People may simply draw upon disparate evidence. Both may be justified; However, if there is a specific proposition expressed by (p) which one accepts and the other rejects, at *most* one of them has knowledge.

6 The Theory of Knowledge and Knowledge Representation

The promise at the beginning of this lecture was that there was some interesting connections between the theory of knowledge and knowledge representation. Here, a few proposals come to mind.

1. Getting our terms straight. I would submit that the term 'knowledge' is grossly abused in the KRR literature. It sometimes means 'belief', 'true belief', 'necessary true belief', 'internally-justified belief', 'rational acceptance', 'consistent belief', or simply sets of propositional literals collected in a data structure that is more sophisticated than the structure of a set.

2. Once our terms are straight, we may be in a position to be more specific about the kinds of epistemic model we wish to advance. It may turn out that the proper model for an epistemic agent is not within some specific logic but, rather, a collection of components. Witness the shift in thinking within natural language processing and speech recognition (e.g., TRAINS at Rochester) from single frameworks to "a patchwork" of technologies.

3. Notice the advantage we have as designers when constructing agents. We are often in a position of having a view of the proper behavior we want our agent to have and also some say over how it will represent this view; in other words, we are in a position to test our models against some specified standard. The truth-conduciveness role will presumable be handled (in part) by the calibration of an agent's sensors, which will determine whether the agent reliably converts pixel imagines into higher-level statements like "there is a human face before me". We then may specify the semantic structure for such statements that is then (presumably) manipulated internally by the agent by reasoning faculties that we can recognize as rational.

I don't suggest that I have advanced here a novel idea; rather, I do suggest that epistemology helps us in two respects in the advancement of this project. First, it offers guidance for keeping our terms straight. Second, it provides an intuitive understanding of these terms, and marks our limits in understanding of how these notionsbehave, and how they contribute to the behavior of the fixation and change of epistemic states. Most of the time, our programs are not modeling truth but rational acceptance (justified belief) (Hawthorne and Bovens, 1999; Wheeler, 2005).

References

[1] Bonjour, L. 1985. *The Structure of Empirical Knowledge*, Cambridge, Mass.: Harvard University Press.

[2] Bonjour, L. 2002. *Epistemology*, Oxford: Rowman and Littlefield.

[3] Bonjour, L. and E. Sosa. 2003. *Epistemic Justification*, Oxford: Blackwell.

[4] Chisholm, R. 1989 (1966). *Theory of Knowledge*, 3rd edition, Englewood Cliffs, NJ: Prentice-Hall.

[5] Churchland, P. M. 1981. "Eliminative materialism and the propositional attitudes", *Journal of Philosophy*, 88(2): 67-90.

[6] Cohen, S. and K. Lehrer. 1983. "Justification, truth and coherence", *Synthese*, 55.

[7] De Rose, K. and T. Warfield. 1999. *Skepticism: A Contemporary Reader*, Oxford: Oxford University Press.

[8] Feldman, R. 1974. "An Alledged Defect in Gettier Counterexamples", *Australasian Journal of Philosophy* (52): 68-69.

[9] Foley, R. 1987. *The Theory of Epistemic Rationality*, Cambridge, Mass.: Harvard University Press.

[10] Fumerton, R. 1990. "The internalism/externalism controversy", in James Tomberlin (ed.), *Philosophical Perspectives, 2: Epistemology.*

[11] Gettier, E. 1963. "Is Justified True Belief Knowledge?" *Analysis* 23 (6): 121-123.

[12] Goldman, A. 1967. "A Causal Theory of Knowing", *Journal of Philosophy* 64: 357-372.

[13] Goldman, A. 1986. *Epistemology and Cognition*, Cambridge: Harvard University Press.

[14] Harman, G. 2001. "Internal Critique: A Logic is not a Theory of Reasoning and a Theory of Reasoning is not a Logic", appearing in *Studies in Logic and Practical Reasoning, Vol. 1.* Gabbay, D. *et. al.* [ed.]. London: Elsevier Science.

[15] Hawthorne, J. and L. Bovens. 1999. "The Preface, the Lottery, and the Logic of Belief", *Mind* 108: 241-264.

[16] Hintikka, J. 1962. *Knowledge and Belief: An Introduction to the Logic of the Two Notions*, Ithaca, NY: Cornell University Press.

[17] Kim, J. 1988. "What is 'Naturalized Epistemology'?" *Philosophical Perspectives 2*, J. Tomberlin [ed.]. Atascadero, CA: Ridgeview Publishing, 381-406.

[18] Kyburg, H. E., Jr. 1961. *Probability and the Logic of Rational Belief.* Middletown, CT: Wesleyan.

[19] Kyburg, H. E., Jr. and C. M. Teng. 2001. *Uncertain Inference*, Cambridge: Cambridge University Press.

[20] Prior, J. 2001. "Highlights of Recent Epistemology", *The British Journal for the Philosophy of Science* 52: 1-30.

[21] Quine, W. V. 1969. *Ontological Relativity and Other Essays*, New York: Columbia University Press.

[22] Sartwell, C. 1992. "Why knowledge is merely true belief", *Journal of Philosophy*, 89: 167-180.

[23] Shope, R. 1983. *The Analysis of Knowing*, Princeton, New Jersey: Princeton University Press

[24] Soames, S. 1999. *Understanding Truth*, Oxford: Oxford Press.

[25] Wheeler, G. 2005. "On the Structure of Rational Acceptance: Comments on Hawthorne and Bovens", *Synthese*, 144(2): 287-304.