



Evidence that stakes don't matter for evidence

Mark Phelan

To cite this article: Mark Phelan (2014) Evidence that stakes don't matter for evidence, *Philosophical Psychology*, 27:4, 488-512, DOI: [10.1080/09515089.2012.733363](https://doi.org/10.1080/09515089.2012.733363)

To link to this article: <https://doi.org/10.1080/09515089.2012.733363>



Published online: 17 Jan 2013.



Submit your article to this journal [↗](#)



Article views: 229



View related articles [↗](#)



View Crossmark data [↗](#)

Evidence that stakes don't matter for evidence

Mark Phelan

Some philosophers have recently defended anti-intellectualism with respect to knowledge and evidence. In this paper, I assess anti-intellectualism about evidence, which claims a relation between one's evidence and the practical benefits or costs of being right or wrong about the propositions supported by that evidence. Proponents of anti-intellectualism generally regard their view as not at all obvious, but nonetheless strongly supported by appeal to our intuitive judgments about whether particular epistemic properties are instantiated in hypothetical cases. Anti-Intellectualism is thus taken by its proponents to be a surprising truth. I show that, though peoples' explicit judgments about the general issue of whether or not non-epistemic factors make an epistemic difference are often in line with anti-intellectualism, their judgments about whether particular epistemic properties are instantiated in hypothetical cases do not display a pattern that would clearly support anti-intellectualism about evidence. Thus, anti-intellectualism about evidence is not entirely surprising, and intuitive assessments of hypothetical cases do not clearly support its truth.

Keywords: Epistemology; Evidence; Experimental Philosophy; K = E; Knowledge; Philosophical Method

1. Introduction

In his influential paper, "Contextualism and knowledge attributions," Keith DeRose asks his readers to consider the following pair of cases:

Bank Case A. My wife and I are driving home on a Friday afternoon. We plan to stop at the bank on the way home to deposit our paychecks. But as we drive past the bank, we notice that the lines inside are very long, as they often are on Friday afternoons. Although we generally like to deposit our paychecks as soon as possible, it is not especially important in this case that they be deposited right away, so I

Mark Phelan is Assistant Professor at Lawrence University.

Correspondence to: Mark Phelan, Department of Philosophy, Lawrence University, Appleton, WI 54911, New Haven, CT 06520-8306, USA. Email: mark.phelan@lawrence.edu

suggest that we drive straight home and deposit our paychecks on Saturday morning. My wife says, “maybe the bank won’t be open tomorrow. Lots of banks are closed on Saturdays.” I reply, “no, I know it’ll be open. I was just there two weeks ago on Saturday. It’s open until noon.”

Bank Case B. My wife and I drive past the bank on a Friday afternoon, as in Case A, and notice the long lines. I again suggest that we deposit our paychecks on Saturday morning, explaining that I was at the bank on Saturday morning only two weeks ago and discovered that it was open until noon. But in this case, we have just written a very large and very important check. If our paychecks are not deposited into our checking account before Monday morning, the important check we wrote will bounce, leaving us in a *very* bad situation. And, of course, the bank is not open on Sunday. My wife reminds me of these facts. She then says, “banks do change their hours. Do you know the bank will be open tomorrow?” Remaining as confident as I was before that the bank will be open then, still, I reply, “well, no. I’d better go in and make sure.” (DeRose, 1992, p. 913)

After posing these cases, DeRose asks us to “assume that in both cases the bank will be open on Saturday and that there is nothing unusual about either case that has not been included in” the description just given (DeRose, 1992, p. 914). DeRose then claims that, when the protagonist (call him “Keith”) says, in *Bank Case A*, that he knows that the bank will be open on Saturday, what he says *seems to be true*. DeRose also claims that, when Keith says, in *Bank Case B*, that he does not know that the bank will be open on Saturday, what he says again *seems to be true*. But how could Keith be truthfully claiming to know in case A, but truthfully claiming not to know in case B, given all the similarities between the two cases?

DeRose answers this question by claiming that the truth conditions of sentences of the form “S knows that p” are context-sensitive. When Keith attributes knowledge to himself in case A, the context in which he makes that attribution is one in which ‘know’ takes on a broad extension: lots of possible states count as “knowing” in the sense in which Keith uses that term in case A. But when Keith denies that he knows in case B, the context in which he makes that denial is a context in which ‘know’ has a much narrower extension: far fewer possible states count as “knowing” in the sense in which the term is used in case B.

That, at any rate, is DeRose’s explanation of what’s going on in the bank cases. But some philosophers have favored an alternative explanation of how Keith could truthfully be claiming to know in *Bank Case A* while also truthfully denying that he knows in *Bank Case B*. According to this alternative explanation, it is not that the extension of ‘knows’ changes across the two cases. Rather, Keith falls into that extension in *Bank Case A*, but fails to fall into that same extension in *Bank Case B*, and that is because *the cost to him of being wrong* about the bank’s being open is higher in *Bank Case B* than it is in *Bank Case A*.

More generally, these philosophers defend the following view.

Anti-Intellectualism about Knowledge (AIK): whether S knows that p at a particular time t constitutively depends upon the costs of S’s being wrong about p at t. The higher those costs, the more stringent are the conditions required for S to know that p.¹

Now, even the proponents of AIK do not regard AIK itself as an obvious truth, but rather as a truth that can be established only by reflecting on our intuitive judgments about cases like, say, DeRose's bank cases. Thus, the first words on the back cover of Stanley's (2005) monograph *Knowledge and practical interests* are these: "Jason Stanley presents a startling and provocative claim about knowledge: that whether or not someone knows a proposition at a given time is in part determined by his or her practical interests." And after defining intellectualism as the view that "knowledge does not depend upon practical facts," Stanley writes that "intellectualism is a wide orthodoxy" (2005, p. 6). Stanley's challenge to intellectualism—his defense of AIK—is based on his acceptance of the principle that, to put it roughly, one may act only on what one knows.

But his acceptance of this principle is in turn based, at least in part, on what he takes to be our intuitive judgments about cases. For instance, after describing what he takes to be our intuitive judgments about five possible cases, Stanley writes: "the intuitions therefore provide powerful intuitive evidence for an antecedently plausible principle concerning the relation between knowledge and action" (2005, p. 11). For Stanley, then, although AIK is not an obvious truth, it can be defended, and at least part of this defense involves appeal to our intuitions about cases.

Given this recent history, and given the tight connection that many epistemologists take there to be between knowledge and evidence,² it will come as no surprise that at least one philosopher (Stanley, 2005) has proposed another view distinct from, but closely related to, AIK.³ In its purest form, this view—Anti-Intellectualism about Evidence (AIE)—claims a relation between one's evidence and the practical benefits or costs of being right or wrong about the propositions supported by that evidence. AIE can be stated a little more precisely in each of two ways, depending on one's view of evidence. We can state AIE as a relation between practical costs and *the constituents of one's evidence set*, as follows:

Anti-Intellectualism about Evidence-Set (AIE-S): whether p is an element of X 's evidence set at a particular time t constitutively depends upon the costs of X 's being wrong about q , a proposition supported by p , at t . The higher those costs, the more stringent are the conditions required for p to be an element of X 's evidence set at t .⁴

AIE-S is a consequence of AIK and the currently popular thesis that one's total evidence is all and only what one knows ($E = K$).

Alternatively, we can formulate AIE, as Stanley does, as a relation between practical costs and *the quality of one's evidence*:

Anti-Intellectualism about Evidence-Quality (AIE-Q): "the quality of evidence for a person X at a time t provided by some information, or some method of gaining information, depends upon whether it is used to support X 's belief in a proposition that is a serious practical question for X at t " (Stanley, 2007, p. 201).

What is it for a proposition to be a "serious practical question" for an individual at a time? Here's what Stanley tells us:

Consider the proposition that you have an even number of hairs. I do not care how many hairs you have, but I believe that you have an even number of hairs. If you

turned out to have an odd number of hairs, that would make no difference to me at all. The various options at my disposal are to retain the belief that you have an even number of hairs or give it up. Given that I do not care about the number of hairs you have, whether or not you have an odd number of hairs will not make a difference to the warranted expected utilities of retaining or discarding my belief. So, the proposition that you have an odd number of hairs is not a serious practical question for me. (Stanley, 2005, p. 96)

It seems, then, that for a proposition to be a serious practical question for an individual at a time is for the truth value of that proposition to make a difference to the warranted expected utilities of believing the proposition.

As we shall see later on, it matters which specific version of AIE one embraces. But just as AIK is not an obvious truth, and requires (and has been alleged to receive) some support from our intuitions about cases, presumably any defense of either version of AIE would require some support from our intuitions about cases. And even if the proponent of AIE could find a plausible argument for some version of AIE that did not rest on our intuitions about cases, it would still impose an explanatory burden on him or her if our intuitions about cases were not consistent with AIE, for he or she would then have to explain away these intuitions. So it should be a question of some interest to the proponent of AIE whether our intuitions about cases are consistent with AIE. While much intuition thumping has taken place with respect to AIK, not nearly as much attention has been paid to AIE. So the question that I set out to answer was the following: do our intuitive judgments about cases support the hypothesis that there is a relation between one's evidence and the practical costs of being wrong?

It may seem at once that there is a problem trying to gather intuitive judgments that bear on either version of AIE. Non-Philosophers will comfortably speak of the scientific evidence in favor of global warming, or the evidence of an oncoming recession, or the evidence of a defendant's guilt. But there's no guarantee that the term "evidence" is ordinarily used in as broad a way as epistemologists use it. For example, some epistemologists speak of an agent's *evidence set*, i.e., the whole body of his or her evidence.⁵ This is not just his or her evidence for or against some particular proposition, but rather his or her evidence for anything whatsoever. That concept does not seem to enjoy comfortable, widespread, pre-theoretical use. And though in ordinary discourse we may be more comfortable speaking in terms of the "quality of one's evidence" for some proposition or other, the way we ordinarily use this phrase may differ from the use made of it by the proponent of AIE-Q. So how can we gather intuitive judgments about whether or not a particular proposition, person, time triple falls into the extension of the theoretical concept of evidence?

My procedure is to locate a point of contact between ordinary terms of epistemic appraisal, on the one hand, and the epistemologist's concept of a person's evidence, on the other. I assume that one point of contact, which will hold for proponents of AIE-S and AIE-Q, is between the epistemic notion of evidence, on the one hand, and plausible norms for confidence, on the other. Specifically, I assume that the more or

better one's evidence for a proposition, the greater one's confidence in that proposition *ought* to be, and vice versa.

I'll put this point as follows:

Bridge from Rational Confidence to Evidence (BRCE): peoples' implicit commitments about an agent's evidence set *or* quality of evidence are reflected in their explicit intuitive judgments about how confident that agent ought to be in various propositions supported by that evidence.

Of course, some readers might reject BRCE.⁶ If these readers can suggest other points of contact between the epistemic notion of a person's evidence set and ordinary terms of epistemic appraisal, it would be helpful to make these known. Such points of contact would provide other opportunities for assessing AIE. Without *some* such points of contact, however, it is unclear why we should find the epistemologist's notion of evidence at all interesting.

Let me briefly consider one other objection to a project of empirically testing the intuitions purported to support AIE. Some may object because they conceive of AIE not as a descriptive hypothesis but rather as a normative epistemological doctrine. What would it mean for an empirical study such as mine if AIE were construed as the theory that evidence *should be* (rather than that it *is*) apportioned relative to costs and practical interests? An epistemologist who construed AIE in this way might eschew people's implicit commitments about evidence in making the case for his or her normative theory, relying instead on other considerations, such as the expected utility of epistemic policies or the inherent dignity of epistemic agents. An epistemologist *might* rely on such considerations. But given the previously discussed importance epistemologists have placed on intuitive judgments about cases, we are presented with the following dilemma: either the prominent epistemologists cited above have intended to establish a descriptive version of AIE (and related epistemic doctrines), for which people's implicit commitments about evidence are clearly relevant, or they have deemed people's implicit commitments about evidence relevant to their particular arguments for a normative version of AIE. Either way, implicit commitments about evidence deserve attention.

Against the background of these assumptions, I conducted several experiments to assess whether our intuitive judgments about cases favor AIE in either of its versions. Does one's evidence at a time constitutively depend upon the cost of being wrong about propositions supported by that evidence?⁷ When I originally undertook this project, no experimental assessment of the philosophical cases used to establish a relationship between stakes and epistemic features existed. Since the completion of the first draft of this paper though, several other papers related to this topic have appeared and found quick publication.⁸ However, while these papers—Buckwalter (2010); Feltz and Zarpatine (2010); May, Sinnott-Armstrong, Hull, and Zimmerman (2010); Pinillos (2012); and Schaffer and Knobe (2010)⁹—address the general topic of stakes sensitivity in epistemology, my research is distinctive in two important respects.

First, all of these previous papers address the stakes sensitivity of *knowledge* (or the truth conditions of *knowledge claims*), whereas the current paper addresses the related

issue of whether *evidence* is sensitive to practical costs. Of course, if, as Williamson maintains, “knowledge, and only knowledge, constitutes evidence” (2000, p. 185), that is no distinction at all. But the case for $E=K$ is, at best, inconclusive. For example, as Comesaña and Kantin (2010) argue, $E=K$ is inconsistent with the existence of Gettier cases. Specifically, if you are justified in believing that whoever got the job has ten coins in his pocket in virtue of your belief that Jones got the job, then your belief that Jones got the job is part of your evidence for believing that whoever got the job has ten coins in his pocket. However, in the famous Gettier case, you do not know that Jones got the job, since that is merely a false belief.¹⁰

The present paper can thus be seen as making a contribution to the unsettled debate over $E=K$. If assessments of evidence diverge from knowledge assessments in systematic ways, this could be interpreted as a strike against $E=K$. On the other hand, accordance between the former and the latter could be seen as evidence in favor of $E=K$, or, alternatively, used to motivate a potential explanation for why proponents of $E=K$ have found the view attractive. Second, whereas previous theorists have merely assessed whether folk intuitions actually support the stakes sensitivity of knowledge, the present paper not only attempts a similar assessment but also offers a diagnosis of the philosophical support for anti-intellectualism.

2. First Experiment: Testing Anti-Intellectualism about Evidence

If anti-intellectualism about evidence were correct, then (given BRCE) we would expect people to judge that an agent should be more confident about p in a case in which the costs to that agent of being wrong about p are low (a “low stakes case”) than in a case in which the costs to the agent of being wrong about p are high (a “high stakes case”).

With this expected asymmetry in mind, in my first experiment I tested participants’ reactions to three conditions. Participants in one condition were asked to assess how confident an agent should be in a high stakes case. In another, they assessed a low stakes case. Participants in a third condition were asked to assess both cases. Thus, the third condition most closely resembled the situation of those reading papers propounding anti-intellectualism. Those papers, following DeRose (1992), invite readers to assess minimally divergent case pairs. I asked 107 University of North Carolina, Chapel Hill undergraduate students in introductory philosophy classes about variations on the following vignette (sentences which vary between vignettes are underlined). The mean number of previous philosophy classes taken by participants in this study was 1 (the mode was 0). No epistemology classes were surveyed in any of the studies discussed in this paper:

Unimportant (Passerby): Kate is ambling down the street, out on a walk for no particular reason and with no particular place to go. She comes to an intersection and asks a passerby the name of the street. “Main Street,” the passerby says. Kate looks at her watch, and it reads 11:45 AM. Kate’s eyesight is perfectly normal, and she sees her watch clearly. Kate’s hearing is perfectly normal, and she hears the passerby quite well. She has no special reason to believe that the passerby is

inaccurate. She also has no special reason to believe that her watch is inaccurate. Kate could gather further evidence that she is on Main Street (she could, for instance, find a map), but she doesn't do so, since, on the basis of what the passerby tells her, she already thinks that she is on Main Street.

In a second condition, which I will call *Important*, participants were asked about a vignette identical to *Unimportant* except that the first sentence was changed to, "Kate needs to get to Main Street by noon: her life depends upon it."¹¹ And in a third condition—*Two Cases*—participants were asked about both the unimportant and important vignettes. For each vignette received, each participant was asked, "how confident should Kate be that she is on Main Street?" This question was judged on a seven-point scale, with 1 representing 'not confident' and 7 representing 'very confident'.¹²

Given the difference in what is at stake, if anti-intellectualism about evidence were correct, we might expect a significantly higher rating of confidence in *Unimportant* than in *Important*.¹³ However, this is not the result that was found. There was no significant difference between participants' average rating for *Unimportant* (5.11) and their average rating for *Important* (4.93).¹⁴

Assuming my assessment was powerful enough to detect a difference if there really was one—an assumption I will discuss later in this paper—these results are unexpected from the perspective of AIE-Q. According to that specification of AIE, the quality of an agent's evidence depends upon whether the evidence is used to support belief in a proposition that is a serious practical question for the agent. In *Important*, as opposed to *Unimportant*, the relevant information is so used. So the quality of the evidence in *Important* should be deemed worse, and that should be reflected in significantly lower confidence ratings in *Important* as opposed to *Unimportant*.

However, a slight caveat arises concerning the bearing of this evidence on AIE-S. According to that specification of AIE, whether a potential piece of information is part of one's evidence set constitutively depends upon the costs of being wrong about that evidence. But, of course, that isn't *all* that membership in an evidence set depends upon. Presumably, other considerations go into this assessment, such as the reliability of the source. If a source is very reliable, the potential information provided by that source may count as evidence *regardless* of how important it is to be right about that information, and vice versa. A proponent of AIE-S may, thus, suggest that a passerby is such a reliable source of information about street names that considerations of importance on the scale the present vignettes offer are insufficient to dislodge potential information from the agent's evidence set. Alternatively, a proponent of AIE-S may contend that a passerby is so unreliable that the potential information she supplies should never enter one's evidence set. I do not find either of these responses particularly plausible.

Nonetheless, these are possible responses that I wish to flag at this point in the dialectic. In following sections, I will discuss more evidence from vignettes involving more and less reliable sources of evidence that makes each of these responses unattractive. These results suggest that across a spectrum of cases, involving highly

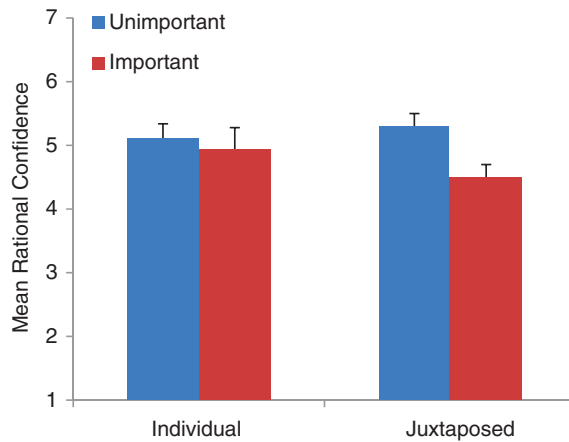


Figure 1. Mean Rational Confidence Ratings for Individual and Juxtaposed Passerby Cases.

reliable, moderately reliable, and highly unreliable sources of information, whether a potential piece of information is thought to be part of one's evidence set does not depend, to any significant extent, upon the costs of being wrong about that information.

If there is no significant difference between participants' judgments about *Unimportant* and *Important*, what explains the motivation for AIE and the significance that epistemologists attach to the topic of anti-intellectualism? To discover this, I looked to *Two Cases*. Here, in a manner familiar from the philosophy papers discussed above, participants were asked to rate Kate's rational confidence in both the unimportant *and* important vignettes. The accuracy of the information in the important vignette was a matter of life and death for Kate, but nothing hinged on the accuracy of the information in the unimportant vignette. Surprisingly, when otherwise equivalent important and unimportant cases were juxtaposed within *Two Cases*, participants responded with significantly distinct answers to the confidence questions. There was a significant difference between participants' confidence ratings for the unimportant case (5.32) and the important case (4.5).¹⁵ The order of these vignettes was counterbalanced between participants, and no order effect emerged.

Let us reflect on the results of the first experiment, represented in Figure 1. While there was a non-significant trend for the individual cases of *Important* and *Unimportant*, this slight difference was dwarfed by the significant difference between participants' confidence judgments for the juxtaposed important and unimportant cases.

In fact, the difference between mean confidence ratings in *Two Cases* was *more than 4½ times* the difference between mean confidence ratings for the non-juxtaposed cases. The difference between effect sizes (reported in the footnotes) was even larger. The effect of importance in the juxtaposed cases was *more than 5 times* that for the non-juxtaposed cases. Thus the first study suggests that while people do not in general make different judgments about an agent's evidence on the basis of

importance when they encounter a single case, they do treat juxtaposed cases quite differently. Or, more accurately, *some people* treat juxtaposed cases quite differently. For this first experiment, 48% responded to the juxtaposed cases in a way consistent with the hypothesis that a piece of information's status or quality as evidence constitutively depends upon the costs of being wrong.

To summarize, in my first experiment I asked participants about one or another of two cases, one in which it was very important that an agent be right about a certain piece of information and another in which it was significantly less important. Assuming BRCE, if either version of anti-intellectualism about evidence were reflected in ordinary intuitions about cases, we would expect a difference between participants' confidence ratings regarding these two cases. However, there was no such difference. Surprisingly, when I juxtaposed cases and asked participants how confident an agent should be, a significant difference did emerge. What explains the difference between participants' responses to juxtaposed cases and their responses to non-juxtaposed cases? One might claim that the stakes stipulated in each case become salient enough to be noticed by participants only when the cases are juxtaposed. Alternatively, one might suggest that juxtaposing the cases gives participants guidance as to how to use the provided scale to rate levels of rational confidence (they can rate it a little lower in the one case than in the other). I will argue that neither of these putative explanations are correct.

My hypothesis regarding AIE, which I will test in the remainder of this paper, has two parts: (1) importance does not, in general, factor into people's assessments of an agent's evidence in particular cases. But (2) some people accept the *general* claim that importance does factor into what evidence someone has; and, when asked about juxtaposed cases, these people tend to adjust for importance when making evidential assessments, in light of their acceptance of this general claim. If this hypothesis is correct, then the fact that some people accept that importance factors into what evidence someone has constitutes evidence in favor of AIE. However, this evidence in favor of AIE is mitigated by two pieces of evidence against AIE. First, importance does not factor into people's assessments of evidence in isolated cases. And, second, most people reject AIE even in juxtaposed cases (or when asked explicitly). Since the philosophical contrast cases that support anti-intellectualism are always presented as juxtaposed cases, this hypothesis, if correct, explains why some peoples' intuitions—including some philosophers' intuitions—*seem* to favor anti-intellectualism, despite the fact that anti-intellectualism about evidence is not clearly supported by those intuitions.

3. Second Experiment: Further Evidence

One should never put too much weight on a single experiment. So, I attempted to test the intuitive case for AIE using different vignettes. In a second experiment, 132 participants on Amazon's Mechanical Turk were asked about randomly

assigned vignettes in which a drunk, not a passerby, was the source of information:

Unimportant (Drunk): Kate is ambling down the street, out on a walk for no particular reason and with no particular place to go. She comes to an intersection. A drunk is standing on the corner, looking a bit unsteady on his feet. Kate asks the drunk the name of the street. “Main Street,” the drunk says. Kate looks at her watch, and it reads 11:45 AM. Kate’s eyesight is perfectly normal, and she sees her watch clearly. Kate’s hearing is perfectly normal, and she hears the drunk quite well. She has no special reason to believe that the drunk is inaccurate. She also has no special reason to believe that her watch is inaccurate. Kate could gather further evidence that she is on Main Street (she could, for instance, find a map), but she doesn’t do so, since, on the basis of what the drunk tells her, she already thinks that she is on Main Street.

As in the first experiment, participants in a second condition, *Important*, were asked about a vignette identical to *Unimportant* except that the first sentence was changed to, “Kate needs to get to Main Street by noon: her life depends upon it.”¹⁶ Again, participants in the third condition, *Two Cases*, received both the *Important* and *Unimportant* vignettes. As before, the order of these vignettes was alternated. For each vignette they received, participants were asked the same question as before: “how confident should Kate be that she is on Main Street?” Again, this question was judged on a seven-point scale.¹⁷ Participants ranged in age from 18 to 66, with 80.3% being 35 years of age or younger. 63.6% of participants self-identified as female. 97% of participants identified themselves as native English language speakers (one participant did not identify a first language, the other three participants identified Chinese, Filipino, and Japanese, respectively). 19 participants were rejected from the following analyses for failing a comprehension check.

Once again, as Figure 2 reveals, no significant difference emerged between participants’ average ratings for *Unimportant* (4.29) and *Important* (4.08).¹⁸ However, when these cases were juxtaposed in *Two Cases*, a significant difference

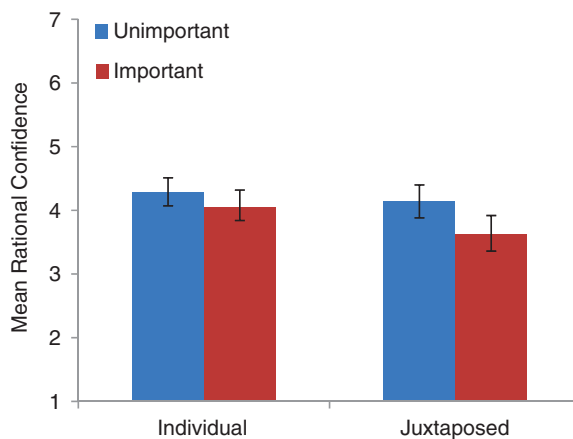


Figure 2. Mean Rational Confidence Ratings for Individual and Juxtaposed Drunk Cases.

emerged between participants' ratings for the unimportant (4.14) and important (3.64) vignettes.¹⁹ For *Two Cases*, 39% of people made judgments consistent with anti-intellectualism.²⁰

The difference between mean confidence ratings for juxtaposed, important, and unimportant cases in the second experiment was not as great as that between juxtaposed important and unimportant cases in the first experiment. But it was still *2½ times* greater than the difference for non-juxtaposed cases in the second experiment. The effect size for the juxtaposed cases was two times greater than that for the non-juxtaposed cases.

For at least a class of vignettes, when high- and low-stakes cases are juxtaposed, a minority of participants are inclined to react differently to prompts about rational confidence than they would react if presented with only one of the prompts; they assess these juxtaposed cases in a way consistent with AIE. On the other hand, when presented with only one non-juxtaposed prompt, participants' confidence assessments are not sensitive to practical stakes. These results (together with BRCE) lend some support to the first part of my hypothesis: importance does not generally factor into peoples' assessments of evidence. But, there are also some objections to the argument thus far that I need to address.

4. Third Experiment: Addressing Difficulties

I envision two main objections to the discussion so far. First, some readers may protest that my method of assessing participants' ratings of rational confidence is flawed. Perhaps, when confronted with only a single case, participants aren't sure how to rate rational confidence using the seven-point scale that I give them, and so their ratings of rational confidence land, more or less arbitrarily, somewhere in the middle of the scale. But when they are given two cases to compare to each other, participants feel constrained to rate the level of rational confidence higher in the low-stakes case than in the high-stakes case, and this comparison between the two cases gives them a bit more guidance concerning how to rate rational confidence in the cases I give them using the seven-point scale. Thus, participants' ratings of rational confidence in the juxtaposed cases are more strongly indicative of their actual assessments of rational confidence than their ratings in the non-juxtaposed cases.

A second objection to the foregoing discussion is as follows: perhaps the reason that participants in my experiments do not rate rational confidence higher for a single low-stakes case than for a single high-stakes case is due simply to the fact that participants didn't really notice, or pay much attention to, the stakes when reading the vignette. Juxtaposing the contrasting cases, however, draws the reader's attention to the stakes, and so readers notice the stakes stipulated in each case. Thus, the objection goes, we should trust ratings of rational confidence in the juxtaposed cases more than assessments of rational confidence in the non-juxtaposed cases.

These objections both claim that, when confronted with a single case, participants' ratings of rational confidence will not properly reflect the extent to which

stakes matter. But, if participants' responses to a single case do not properly reflect the extent to which stakes matter, then they should also not properly reflect the extent to which other, equally salient, factors matter.²¹ I decided to test whether this is the case with respect to a factor that we can all, presumably, agree is relevant to a participants' judgments concerning rational confidence: namely, the reliability of the source of information given in the vignette. I assume that all readers will agree that one should have more confidence in information imparted by a reliable source than in information imparted by an unreliable source. I also take it as uncontroversial that the average passerby is a more reliable source of information than a drunk. A street sign would be an even more reliable source of information about street names than the average passerby.

To determine whether differences of reliability are reflected even in single cases, I first ran another experiment on the model of the above experiments, but, instead of asking about drunks or passersby, I made street signs the source of information for the 141 University of North Carolina, Chapel Hill undergraduate participants. The mean number of previous philosophy classes taken by participants in this study was 1 (the mode was 0). Once again, there was no significant difference between *Unimportant* (5.61) and *Important* (5.79).²² Once again, there was a significant difference for the unimportant case (5.76) and the important case (5.26) in *Two Cases*.²³ Similarly to previous experiments, only a portion of people—36%—judged as anti-intellectualists in the juxtaposed *Two Cases* condition.²⁴

With the results of this third experiment, we are able to compare participants' responses in non-juxtaposed cases involving a highly reliable, a moderately reliable, and a relatively unreliable source of information about street names. In fact, the mean value of participants' answers for the non-juxtaposed cases involving the highly reliable street sign ($M = 5.7$, $SD = 1.16$) was higher than that for cases involving the moderately reliable passerby ($M = 5.02$, $SD = 1.52$), which was higher than that for the unreliable drunks ($M = 4.19$, $SD = 1.4$). Comparing this to the average rating for

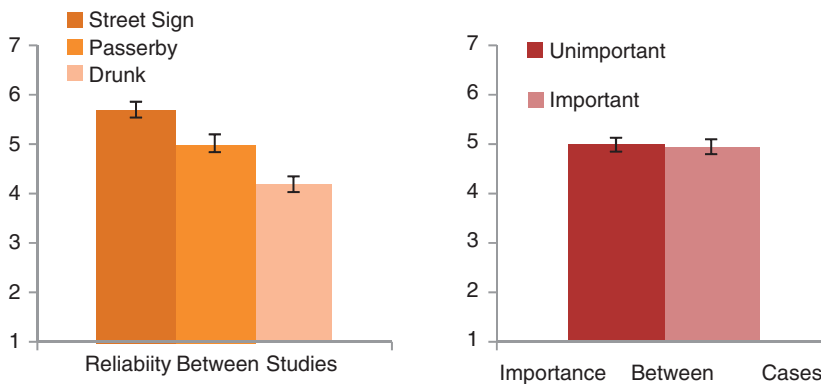


Figure 3. Comparing the Effect of Reliability Between Studies to the Effect of Importance Between Cases.

all non-juxtaposed important ($M = 4.95$, $SD = 1.57$) and unimportant ($M = 4.99$, $SD = 1.42$) cases, we get the following result reflected in Figure 3.

I compared the results for all of the non-juxtaposed cases in the three previous experiments. I ran a 3 (reliability) \times 2 (importance) analysis of variance to determine to what degree the results for each vignette were explained by the reliability of the source, and to what degree they were explained by the practical importance of being right. The dependent variable, confidence ratings, was normally distributed for the groups formed by the combination of the levels of reliability and importance, as assessed using Q-Q Plots. There was homogeneity of variance between groups as assessed by Levene's test for equality of error variances. There was a significant effect of reliability.²⁵ However, no significant effect emerged for importance.²⁶ There was no significant interaction effect. Post hoc comparisons using the Tukey HSD test indicated that the mean confidence rating for non-juxtaposed street sign cases was significantly different than the mean confidence rating for non-juxtaposed passerby and drunk cases. The mean confidence rating for non-juxtaposed passerby cases was significantly different than the mean confidence rating for non-juxtaposed drunk cases.²⁷

Contra the first objection above, participants are not at a loss when rating rational confidence on a seven-point scale for a single case. Confidence assessments are sensitive to the reliability of the source of the information under consideration, and participants' ratings reflect this sensitivity, even when made about a non-juxtaposed case on a single scale in isolation. Furthermore, given that participants' ratings are sensitive to source reliability, participants clearly notice (though perhaps unconsciously) the reliability of the source imparting the information even in single cases. If the reliability of the source of information is of equal salience with what is at stake, the fact that participants' judgments accurately reflect the extent to which reliability matters in non-juxtaposed cases suggests that participants' judgments accurately reflect the extent to which stakes matter in non-juxtaposed cases (i.e., not much). If reliability is as salient as stakes, then, in light of the fact that participants clearly notice reliability, it would be ad hoc to claim that participants fail to notice what's at stake, as the second objection above suggests.

Of course, a proponent of AIE might argue that stakes and reliability are not equally salient features of epistemic appraisal. Recently, in response to this paper, Hansen (unpublished manuscript) has put forward a carefully articulated defense of this claim. Hansen points to work by Hsee, Loewenstein, Blount, and Bazerman (1999), demonstrating that features differ in terms of how easy it is to evaluate them in isolation. Summarizing Hsee et al's discussion, Hansen writes:

Whether an attribute is easy or difficult to evaluate, according to Hsee et al. (1999, p. 578), "depends on the type and the amount of information the evaluators have about the attribute." Relevant information includes which value for the attribute would be evaluatively neutral, what the best and worst values for the attribute would be, and "any other information that helps the evaluator map a given value of the attribute onto the evaluation scale." (unpublished manuscript, p. 16)

Hansen argues that whereas we have a relatively high degree of relevant information about how reliable a source is in non-juxtaposed cases, we have relatively little relevant information about the value of what is at stake in those cases. If this is correct it will be more difficult to evaluate the degree to which a piece of information matters than to evaluate how reliable the source of the information is in non-juxtaposed cases. Thus, what is at stake is less salient than the reliability of the information source in non-juxtaposed cases, and the previous response to the second objection collapses.²⁸

But should we accept that we have a relatively low degree of relevant information about what is at stake in non-juxtaposed cases? Hanson's argument turns on the claim that there is no clear upper bound on importance. Whereas there are clear best and worst values when it comes to reliability—a source of information can be completely unreliable or completely reliable—Hansen claims that there is no most important outcome. As he writes about my cases, “certainly whether someone lives or dies is important, but there's always something more important” (unpublished manuscript, p. 18).²⁹ If there is no most important outcome, there is no evaluative neutral, either. Thus, Hansen concludes, given that there is an upper bound for reliability but not for stakes, we have relatively less relevant information when it comes to assessing stakes in non-juxtaposed cases.

Now, I do not think it is obvious that there is no upper bound for importance of what is at stake. But even if there is no such upper bound, it does not follow that, compared to source-reliability, we have less relevant information for assessing stakes in non-juxtaposed cases. Note that Hsee et al. specify that relevant information for assessing an attribute includes “any . . . information that helps the evaluator map a given value of the attribute onto the evaluation scale” (1999, p. 578). Thus, even if there is no upper bound for importance, without having some reason to suppose that reliability and stakes are on a par so far as other relevant information is concerned, we cannot conclude that we have less relevant information for assessing stakes in non-juxtaposed cases. Clearly, other information can trump the absence of a best or worst value for an attribute, for there are many unbounded attributes that are easy to judge in isolation. For example, I may easily conclude that someone is a huge sleaze, though there is no end to sleaziness.

In fact, we have some reason to suppose that reliability and stakes are not on a par so far as other relevant information is concerned. After all, an outcome (or at least the particular outcomes discussed in my cases) more closely approximates a one-to-one relationship with an importance evaluation than does an information source with a reliability evaluation. We need know nothing more than that some (ordinary) person's life depends upon knowing something to accurately conclude that knowing that thing is very important to that person. On the other hand, knowing that, for example, some person is a normal passerby does not in and of itself give us a conclusive reason for assessing that person as a reliable source of information about street names. Even without a special reason to doubt the accuracy of a passerby, we know that passersby vary widely in their reliability concerning street names in a way that the importance of staying alive does not vary for the average person.

I look forward to more discussion about the comparative salience of reliability and stakes, but on the basis of this discussion I provisionally conclude that reliability is of equal salience with what is at stake in my non-juxtaposed cases. Given that participants notice how reliable a source of information is in non-juxtaposed cases, I thus conclude that, contrary to the second objection discussed at the beginning of this section, participants notice what is at stake in my non-juxtaposed cases.

Comparing cases by reliability (highly reliable, moderately reliable, and unreliable) also helps to resolve a previously discussed caveat concerning the bearing of my results on AIE-S. Considerations of importance seem always insufficient to affect whether or not participants regard a potential piece of information as a member of an agent's evidence set, at least when it comes to non-juxtaposed cases. Had they ever been sufficient, we would have expected confidence judgments in at least one of the non-juxtaposed unimportant cases to be higher than those in the relevant non-juxtaposed important case. Of course, a proponent of AIE-S may contend that membership in an evidence set does depend on importance for some subset of cases involving a source of information even less reliable than a pair of drunks or even more reliable than street signs. But if the effect of AIE-S is discernible only for such a small subset of cases, then it is not an empirically interesting hypothesis.³⁰

Considered independently, the results for *Street Signs* also overturn another potential objection to my method and conclusions. In the first two experiments, there were non-significant trends in the direction anti-intellectualists would predict for the non-juxtaposed cases. So some readers may have felt that anti-intellectualism was exerting an effect in these individual cases as well. Perhaps these experiments simply lacked the statistical power to register this general importance effect. However, the results for the third experiment should help alleviate this concern. In this experiment, the non-significant trend for the single cases was in the direction opposite that anti-intellectualism would predict. That is, in the third experiment, rational confidence was rated *higher* in the non-juxtaposed, *high-stakes* case. This shift in direction of the non-significant trend suggests that stakes make no stable difference to evidence in non-juxtaposed cases, not merely a non-significant one.

Furthermore, assuming an effect size for importance similar to that of our aggregated within-participants experiments,³¹ the power of the aggregated between-participants experiments to detect an effect for importance was greater than 0.99. Cohen (1988) recommends achieving a power of 0.8 in order to be confident that one would have detected an effect had there been one. Thus, I conclude that the between-participants experiments were powerful enough to detect an effect of importance on participants' intuitive assessments of cases had there really been one.

Finally, let us consider a challenge to the studies that cannot be resolved through reflection on my experimental results: one may object to the ecological validity of vignette studies in general. In response to participants' assessments of my non-juxtaposed cases, a proponent of AIE might claim that people do not take the life-and-death situation identified in *Important* seriously. If people were *actually* placed in a life-and-death situation, such a proponent might suggest, their rational confidence might *actually* be affected by the high-stakes. Indeed, such a proponent

might attempt to draw support for this hypothesis from psychological experiments in which stakes are raised in a variety of ways, and participants' decision-making behavior is subsequently observed.³² Participants in such conditions typically search for more information or more carefully analyze what information they have before coming to a decision. In other words, these participants behave as though they were less confident in the evidence they have been given than do controls in cases in which the stakes have not been raised. A proponent might take such findings to support AIE and to tell against my findings.

In fact, such findings are orthogonal to my results and not clearly relevant to AIE. Existing psychological studies assess *actual* confidence, whereas my experiments are intended to assess judgments concerning *rational* confidence. To put the point another way, the psychological studies assess how confident participants *are* in high-stakes cases, whereas the present studies examine how confident participants think that agents *should be* in such cases. The present studies ask participants to make a normative assessment—how confident should an agent be—on the assumption of BRCE, that peoples' implicit commitments about an agent's evidence are reflected in their explicit intuitive judgments about how confident that agent ought to be in various propositions supported by that evidence. Intuitively, how confident one ought to be in a given proposition is proportional to the quality and quantity of the evidence one has for that proposition. It is not reasonable to suppose that one's actual confidence is so proportioned. The present studies, by assessing normative judgments about confidence, provide clues to a descriptive account of evidence. On the other hand, the aforementioned studies provide evidence for a descriptive theory of confidence.

Of course, the peripheral nature of the psychological studies does not settle the ecological validity question. Do people take the life-and-death situation identified in *Important* seriously? A reason for thinking they do not could be that locutions of the form, "my life depends on it," are often uttered ironically. But the ironic interpretation of an utterance is usually motivated by a specific ironic form of intonation (Cutler, 1976). (This is why it is often difficult to convey irony via email.) Obviously, this intonational cue is missing in printed vignettes.

Furthermore, participants' responses to the juxtaposed conditions suggest that the phrase, "her life depends upon it," is actually understood by (at least some) participants to raise the stakes. In each experiment, a sizable minority of participants in the juxtaposed condition (41% across all studies' juxtaposed conditions) judged that Kate ought to be less confident that she was on Main Street in the important vignette. Presumably, an explanation for this difference in confidence assessments in juxtaposed cases will invoke a perceived shift in stakes between the two cases. But if people take the life-and-death situation identified in the important vignette seriously when it is presented alongside another vignette, it is reasonable to suppose that they take it seriously when it is presented alone, in *Important*.

Finally, even if ecological validity constituted a successful challenge to my attempted assessment, it would also undermine the philosophical case for AIE. As I discussed above, that case is largely built on intuitive reactions to printed stories, not

upon such reactions to real life-or-death situations. Of course, it would be interesting to know how real high-stakes situations affect rational confidence. But absent such knowledge, we have to rely on the method of cases.

In this section, I have considered five potential problems for the non-juxtaposed studies discussed in this paper. These potential problems respectively held that:

- Participants are at a loss when rating rational confidence on a seven-point scale.
- Participants fail to notice what's at stake when reading non-juxtaposed cases.
- Though considerations of stakes factor into whether a piece of information belongs in one's evidence set, they are swamped by other epistemic factors.
- My non-juxtaposed experiments lack the statistical power to capture the real effect that considerations of practical importance exert on evidence.
- The current vignette studies are not an ecologically valid paradigm to investigate AIE.

I have offered reasons for rejecting each problem. Having addressed these objections, I conclude that my results for the non-juxtaposed cases suggest that importance does not, in general, factor into peoples' assessments of evidence. But what about the juxtaposed cases? I now attempt to test the second part of my hypothesis: when important and unimportant cases are juxtaposed, do some people judge consistently with anti-intellectualism because of their commitment to the general claim that the importance of being right matters to what one's evidence is?

5. Fourth Study: Assessing Our General Views about Evidence

Although people do not generally take importance into account when assessing an agent's evidence in a particular case, importance does influence some participants' judgments when important and unimportant contrast cases are juxtaposed. What might explain this unusual influence of importance? I hypothesize that it is explained by some participants' commitment to a general principle that importance is a factor relevant to assessments of evidence. Juxtaposing important and unimportant cases reminds participants of this commitment, and may thereby trigger this commitment for those participants who have it.

To gather support for the hypothesis that some people have a commitment to the relevance of importance to evidential assessments, I probed participants about their general commitments regarding evidence. 69 University of North Carolina, Chapel Hill undergraduates were presented with the following short vignette:

Kate needs to get to Main Street by noon. She comes to an intersection and asks a passerby the name of the street. The passerby says, "Main Street." Kate looks at her watch, and it reads 11:45 AM. Kate wonders how confident she should really be that she is on Main Street.

Participants were then invited to reflect on and articulate their general commitments about evidence by responding to the question, "what factors *should* affect Kate's confidence that she is on Main Street?" Participants were requested to check all of those that apply for nine different factors ranging from the country in which Kate

lives, to the passerby's emotional state, to the reliability of past information from random people.³³ Participants could also select "how important it is that Kate be on Main Street at noon." In fact, 30 participants, or 43%, selected this importance factor as one that should affect Kate's confidence. Thus, a proportion of people when asked directly avow beliefs consistent with anti-intellectualism about evidence. The percent of participants who judged as anti-intellectualists across all of the previous juxtaposed conditions was 41%. Therefore, a percentage of people hold explicit beliefs consistent with AIE, and a similar percentage judge consistently with AIE when presented with juxtaposed important and unimportant cases. Obviously, this does not confirm the second part of my hypothesis regarding AIE, but it does suggest that the hypothesis merits future investigation. For some people, juxtaposing important and unimportant cases may trigger a commitment to anti-intellectualism.

6. Conclusion

Proponents of anti-intellectualism in epistemology generally regard the view as a surprising truth, which is supported by appeal to our intuitive judgments about whether certain epistemic properties are instantiated in hypothetical cases. The present discussion suggests that anti-intellectualism about evidence is not a surprising truth, for, when asked explicitly, many ordinary people avow a commitment to it. Indeed, this explicit commitment to AIE on the part of a sizable minority of participants may explain anti-intellectualist assessments of juxtaposed cases. If this is correct, then AIE is not a surprising truth supported by our intuitive judgments about hypothetical cases; it is an explicitly avowed commitment held by some people that leads them to assess certain hypothetical contrast cases in predictable ways. At the same time, reflection on the non-juxtaposed cases reveals that when the contrast between high and low stakes is not made explicit, importance does not factor into peoples' assessments of evidence. The results suggest that the same participant who rates rational confidence low in a high stakes case when it is juxtaposed with a low stakes case would not rate rational confidence low were he or she presented with a non-juxtaposed high stakes case.

An analogy may help illustrate the situation for AIE. Consider a proponent of Basenjjs who wants to argue that Basenjjs are better behaved than Boxers. Suppose a sizable minority of people shares this opinion, and hardly anyone explicitly believes that Boxers are better behaved than Basenjjs. When people observe a Basenji and then observe a Boxer, and are subsequently asked to assess which one behaves better, most people conclude that they behave about the same, but a sizable minority of people (about the same number as explicitly believe that Basenjjs behave better) reliably assesses the Basenji as having behaved better than the Boxer. However, whenever anyone observes just a Basenji, they assess it as behaving about as well as they would have assessed a Boxer, had they observed just a Boxer. If this is the case, should we side with the Basenji proponent and conclude that Basenjjs really do behave better than Boxers? Perhaps, though we would need some reason for prioritizing juxtaposed

observations of Basenjis and Boxers over non-juxtaposed observations. And we would also need some justification for supposing that a minority preference for Basenji behavior in juxtaposed observations reflects real discernment of better Basenji behavior, not simply self-affirmation of one's explicit commitments.

We face a similar situation when it comes to the juxtaposed assessments that seem to support AIE. I have argued above that non-juxtaposed observations provide a good measure of the degree to which stakes matter for evidence. It remains for the proponent of AIE to argue why juxtaposed observations are a better measure, and why they should be prioritized over non-juxtaposed observations. Even if the proponent of AIE is able to do this, it remains to be shown that anti-intellectualist assessments of juxtaposed cases are sensitive to the appropriate factors, i.e., those factors that are relevant to the truth or falsity of the stakes' sensitivity of evidence. If anti-intellectualist responses to juxtaposed cases are explained in virtue of an explicit commitment to AIE on the part of some participants, it becomes tough to argue that such responses are sensitive to the appropriate factors without a thorough investigation of this commitment. Where did it come from, and why does it arise in some but not other people?³⁴ In light of these considerations, AIE remains an open empirical question. Though, we now have some evidence that stakes don't matter for evidence.

Acknowledgements

This paper has a long prepublication history, so I am sure I will leave many helpful commentators out of these acknowledgements. Nonetheless, thanks to Joshua Alexander, Nat Hansen, Eric Mandelbaum, Jonathan Weinberg, and two anonymous referees for *Philosophical Psychology*, who all provided helpful comments on earlier drafts of this paper. Thanks also to other members (besides Alexander and Weinberg) of the Experimental Epistemology Lab at the University of Indiana (who read an early version of this paper) and to audiences at the Experimental Epistemology Workshop (in Buffalo, NY in October 2009) and the Interdisciplinary Approaches to Philosophical Issues Conference (in Mobile, AL in September 2008), where I gave invited presentations on this project. A long discussion of an early version of this paper took place on the philosophy weblog, *Certain Doubts*, where many ultimately helpful criticisms and suggestions were raised. I am especially grateful to the help and advice of Joshua Knobe and Ram Neta. The present paper is a descendant of some earlier unpublished work by Phelan and Neta.

Notes

- [1] For an early endorsement of anti-intellectualism concerning knowledge, though not by that name, see Levi (1967). More recently, see Fantl and McGrath (2007), Hawthorne (2004), Kaplan (2006), and Stanley (2005).
- [2] Williamson (2000) famously argues that S knows that p if and only if p is an element of S's evidence set. I follow Williamson in abbreviating this thesis as E = K. This thesis is endorsed by Bird (2005) and Hyman (2006), among others.

- [3] Stanley claims that his “own view is that all epistemic notions are interest-relative” (2005, p. 182), though he does not explicitly defend this sweeping conclusion.
- [4] This version of AIE is suggested by the following passage: “these points lead us to the second and opposing moral that may be drawn from the fact that knowledge is an interest-relative notion. It is *prima facie* difficult to accept that one person knows that *p* and another does not, despite the fact that they have the same evidence for their true belief that *p*. But, if knowledge is anywhere near as central to epistemology as the considerations in Williamson (2000) suggest, then one would expect that evidence is similarly interest-relative” (Stanley, 2005, p. 181). In this passage, Stanley takes AIE to be suggested by AIK and Williamsonian considerations, which are considerations in favor of $E = K$. But the only version of AIE that is implied, or even clearly suggested, by AIK and $E = K$ is AIE-S.
- [5] Other epistemologists (e.g., DeRose, 2000) have expressed doubts about whether there is any such thing as an agent’s *total evidence*, as opposed to his or her evidence *for* one or another particular hypothesis.
- [6] As I discuss in note 30, a leading epistemologist has suggested that the ordinary term ‘confidence’ may be ambiguous, and that BRCE might be true on only one of the term’s readings. Fortunately, careful consideration of results presented later in this paper suggest that participants understand ‘confidence’ in a way that supports BRCE.
- [7] Constitution talk is used in a wide variety of senses within philosophy, so it may help to have a gloss of what I mean by the term. In the case of AIE-S, when I ask whether evidence constitutively depends upon practical costs, I am wondering whether a purported piece of evidence’s *status as part of one’s evidence set* can fluctuate with the importance of knowing propositions supported by that evidence (when other variables are held fixed). When I ask the same question regarding AIE-Q, on the other hand, I am wondering whether *the quality* of a piece of evidence can fluctuate with the importance of knowing propositions supported by that evidence (when other variables are held fixed).
- [8] See Buckwalter (2012) for an insightful discussion of the current state of play, including this paper.
- [9] Two of these (May et al., 2010; Schaffer & Knobe, 2010) cite an earlier version of this paper.
- [10] Joyce (2004) and Goldman (2009) also present important challenges to $E = K$.
- [11] I present the vignettes in this way to maintain the flow of the paper. However, I have included the full vignettes in the appendix.
- [12] A seven-point scale was selected because I wanted to give participants sufficient options to reflect their true feelings about the case, while at the same time allowing a neutral response option.
- [13] Provided, of course, that BRCE is correct, that participants report their intuitive judgments about rational confidence when presented with this question, and that the manipulation was powerful enough to detect a difference. I will assume the first two points, and discuss the third later.
- [14] SD for unimportant = 1.23, SD for important = 1.77, $t(55) = 0.435$, $p = 0.665$ (two tailed), $d = 0.12$.
- [15] SD for unimportant = 1.27, SD for important = 1.39, $t(49) = 4.499$, $p < .001$ (two tailed), $d = 0.62$. I used an alpha level of .05 for this and all other statistical tests reported in this paper.
- [16] This *Important* vignette appears in its entirety in the appendix at the end of this paper.
- [17] This study supersedes a previous study, conducted at University of North Carolina, Chapel Hill, involving two drunks. An anonymous referee for this journal suggested that two drunks could be thought to be a more reliable source of information than a single passerby. As alluded to previously, perceived reliability of an information source will be a crucial factor later in the paper. Thus, I removed the potential confound by conducting a replacement study, replacing the drunks with a single drunk. Study materials differed only in terms of the vignettes. Substantial differences with the previous drunks vignettes were constrained to the

third through fifth sentences, which previously read: “two drunks are standing on the corner in an intense discussion. Kate asks the drunks the name of the street. ‘Main Street’, the drunks say.” As detailed in notes 20 and 27, results for the new drunk cases did not differ significantly from results for the old drunks cases and support all of the same inferences.

- [18] SD for unimportant = 1.35, SD for important = 1.46, $t(72) = 0.63$, $p = 0.531$ (two-tailed), $d = 0.15$.
- [19] SD for unimportant = 1.55, SD for important = 1.79, $t(35) = 2.24$, $p = 0.032$ (two-tailed), $d = 0.30$. (Three additional participants who passed the comprehension check were rejected from this analysis because they did not rate at least one of the vignettes.)
- [20] Results for the drunk cases presented here mirrored results for the cases involving two drunks, which the cases presented here supersede (see note 17). For the original non-juxtaposed drunks cases, scores for *Unimportant* ($M = 4.72$, $SD = 1.23$) did not differ significantly from *Important* ($M = 4.39$, $SD = 1.91$). $t(34) = 0.622$, $p = 0.538$ (two tailed), $d = 0.20$. For the original juxtaposed drunks cases, scores for *Unimportant* ($M = 4.53$, $SD = 1.32$) did not differ significantly from *Important* ($M = 3.69$, $SD = 1.47$). $t(35) = 4.14$, $p < 0.001$ (two-tailed), $d = 0.60$. 42% of participants to the original juxtaposed drunks cases responded as anti-intellectualists.
- [21] In a recent manuscript critiquing this paper, Hansen points out that the contrapositive of this important assumption is easier to understand: “if participants’ responses to a single case properly reflect the extent to which factors that are equally salient to stakes matter, then they should also properly reflect the extent to which stakes matter” (unpublished manuscript, p. 13). The discussion of reliability in the next several paragraphs provides evidence in favor of the antecedent of this conditional. First, experimental results suggest that participants are sensitive to the reliability of the source when assessing evidence. I then offer considerations in favor of the crucial parity claim, that stakes and reliability are equally salient in the non-juxtaposed cases.
- [22] SD for unimportant = 1.34, SD for important = 0.96, $t(73) = 0.675$, $p = 0.502$ (two-tailed), $d = 0.15$.
- [23] SD for unimportant = 1.55, SD for important = 1.44, $t(65) = 2.92$, $p = 0.005$ (two-tailed), $d = 0.34$.
- [24] The vignettes used in this experiment are presented in detail in the appendix.
- [25] $F(2, 204) = 23.07$, $p < .001$, $\eta^2 = 0.19$.
- [26] $F(1, 205) = 0.123$, $p = 0.726$.
- [27] A worry arises for the comparison discussed in the text: responses for some vignettes (street sign, passerby) were generated by college students, whereas responses to others (drunk) were generated online, by workers on Amazon’s Mechanical Turk. As mentioned in note 17, the drunk experiment supersedes a drunks experiment, which was conducted on the same college student participants as street sign and passerby. The drunk experiment controls for a potential confound in the original drunks experiment, but was conducted at a time when the original population of college students was no longer available. I have discussed the drunk experiment in the body of the text, but the drunks experiment would also support all inferences made above. As discussed in note 20, results for the drunks cases mirror the drunk cases, both in juxtaposed and non-juxtaposed conditions. A series of independent samples T-tests also revealed no significant difference between responses to the drunks and drunk cases for non-juxtaposed important vignettes ($t(54) = 1.15$, $p = 0.255$), non-juxtaposed unimportant vignettes ($t(52) = 0.652$, $p = 0.517$), juxtaposed important vignettes ($t(71) = 0.02$, $p = 0.983$), and juxtaposed unimportant vignettes ($t(70) = 1.145$, $p = 0.256$). And, most importantly, a 3 (reliability) \times 2 (importance) analysis of variance conducted on the non-juxtaposed street sign, passerby, and drunks cases also revealed a significant effect of reliability ($F(2,165) = 9.23$, $p < 0.001$, $\eta^2 = 0.10$), and no significant effect for importance ($F(1,166) = 0.044$, $p = 0.835$).

- [28] Hansen recognizes that this conclusion follows only on the assumption that “ease or difficulty of evaluating an attribute is a suitable construal of [my] notion of ‘salience’” (unpublished manuscript, p. 19). I concede that it is.
- [29] Hansen suggests that there does seem to be a lower bound for stakes: “nothing might turn on whether a proposition turns out to be true or false” (unpublished manuscript, p. 18). He concedes that this is reflected in my unimportant cases.
- [30] There is yet another worry about the experimental set-up of these studies that comparing cases by reliability helps resolve. As discussed in reaction to an early version of this paper posted at the weblog, *Certain Doubts*, and raised by Keith DeRose (personal communication, August, 2008), “how confident should Kate be” seems to admit of two uses. Prima facie, the question could be interpreted as asking (as I will argue it is not), “regardless of her actual evidence in this case, how confident should Kate be as a matter of prudence, given that her life depends on being right (or, vice versa, given that nothing much hangs on being right).” Alternatively, the question could be interpreted as asking (as is assumed by BRCE), “how confident should Kate be relative to her evidence,” (where evidence itself—and so participants’ responses—are sensitive to stakes insofar as AIE is correct). Only if the second interpretation is the one participants give to the question do these results bear on AIE. On the first interpretation, the results might reveal a surprising conclusion about what prudential norms of action in high stakes cases the folk adopt, but they would not bear on the question of whether evidence is sensitive to stakes. Fortunately, the comparison of non-juxtaposed cases relative to reliability favors the former interpretation. If participants were interpreting the question as asking about how confident Kate should be given that her life depends (or does not depend) on being right, then we would not expect responses to vary dependent upon the reliability of the source of the information. After all, norms of prudence will depend on what is at stake, not on the reliability of the source of information. Thus, participants’ responses across all manipulations of reliability should average out to about the same mean. However, responses do systematically vary relative to reliability, as people tend to agree that quality of evidence should vary relative to reliability of the source. This suggests that participants understand the question to ask, as the present assessment requires, “how confident should Kate be given her evidence.”
- [31] $t(151) = 5.536$, $p < 0.001$ (two tailed), $d = 0.90$.
- [32] Some such experiments are discussed in Kunda’s (1990) survey. See also Nagel (2008).
- [33] The factors from which participants could choose appear in the appendix.
- [34] Thus, the challenge of the present paper fits into the emerging project of experimental restrictionism. As Alexander writes, this project “puts pressure on our intuition deploying practices. At the very least, philosophers face a dilemma: either we must explain why these kinds of intuitional sensitivity are welcome or we must stop appealing to these philosophical intuitions as evidence and place *local restrictions* on our intuition deploying practices. (The name ‘restrictionism’ derives from the second horn of this dilemma)” (2012, p. 82).

References

- Alexander, J. (2012). *Experimental philosophy: An introduction*. Cambridge, UK: Polity Press.
- Bird, A. (2005). Abductive knowledge and holmesian inference. In T. Gendler & J. Hawthorne (Eds.), *Oxford studies in epistemology* (pp. 1–31). Oxford: Oxford University Press.
- Buckwalter, W. (2010). Knowledge isn’t closed on Saturday: A study in ordinary language. *Review of Philosophy and Psychology*, 1, 395–406.
- Buckwalter, W. (2012). Non-traditional factors in judgments about knowledge. *Philosophy Compass*, 7, 278–289.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York: Academic Press.

- Comesaña, J., & Kantin, H. (2010). Is evidence knowledge? *Philosophy and Phenomenological Research*, 80, 447–454.
- Cutler, A. (1976). Beyond parsing and lexical look-up: An enriched description of auditory sentence comprehension. In R.J. Wales & E. Walker (Eds.), *New approaches to language mechanisms* (pp. 133–149). Amsterdam: North Holland.
- DeRose, K. (1992). Contextualism and knowledge attributions. *Philosophy and Phenomenological Research*, 52, 913–929.
- DeRose, K. (2000). Ought we to follow our evidence? *Philosophy and Phenomenological Research*, 60, 697–706.
- Fantl, J., & McGrath, M. (2007). On pragmatic encroachment in epistemology. *Philosophy and Phenomenological Research*, 75, 558–589.
- Feltz, A., & Zarpentine, C. (2010). Do you know more when it matters less? *Philosophical Psychology*, 23, 683–706.
- Goldman, A. (2009). Williamson on knowledge and evidence. In P. Greenough, D. Pritchard, & T. Williamson (Eds.), *Williamson on knowledge* (pp. 73–91). Oxford: Oxford University Press.
- Hansen, N. (unpublished manuscript). Contrasting cases. (Cited with permission).
- Hawthorne, J. (2004). *Knowledge and lotteries*. Oxford: Oxford University Press.
- Hsee, C.K., Loewenstein, G.F., Blount, S., & Bazerman, M.H. (1999). Preference reversals between joint and separate evaluations of options: A review and theoretical analysis. *Psychological Bulletin*, 125, 576–590.
- Hyman, J. (2006). Knowledge and evidence. *Mind*, 115, 891–916.
- Joyce, J. (2004). Williamson on evidence and knowledge. *Philosophical Books*, 45, 296–305.
- Kaplan, M. (2006). Deciding what you know. In E. Olsson (Ed.), *Knowledge and inquiry: Essays on the pragmatism of Isaac Levi* (pp. 225–240). Cambridge: Cambridge University Press.
- Kunda, Z. (1990). The case for motivated reasoning. *Psychology Bulletin*, 108, 480–498.
- Levi, I. (1967). *Gambling with truth*. New York: Alfred A. Knopf.
- May, J., Sinnott-Armstrong, W., Hull, J.G., & Zimmerman, A. (2010). Practical interests, relevant alternatives, and knowledge attributions: An empirical study. *Review of Philosophy & Psychology*, 1, 265–273.
- Nagel, J. (2008). Knowledge ascriptions and the psychological consequences of changing stakes. *Australasian Journal of Philosophy*, 86, 279–294.
- Pinillos, Á. (2012). Knowledge, experiments, and practical interests. In J. Brown & M. Gerken (Eds.), *Knowledge ascriptions* (pp. 192–219). Oxford: Clarendon.
- Schaffer, J., & Knobe, J. (2010). Contrastive knowledge surveyed. *Noûs*. Advance online publication. doi:10.1111/j.1468-0068.2010.00795.
- Stanley, J. (2005). *Knowledge and practical interests*. Oxford: Oxford University Press.
- Stanley, J. (2007). Replies to Gilbert Harman, Ram Neta, and Stephen Schiffer. *Philosophy and Phenomenological Research*, 75, 196–210.
- Williamson, T. (2000). *Knowledge and its limits*. Oxford: Oxford University Press.

Appendix

Unimportant (Passerby):

Kate is ambling down the street, out on a walk for no particular reason and with no particular place to go. She comes to an intersection and asks a passerby the name of the street. “Main Street,” the passerby says. Kate looks at her watch, and it reads 11:45 AM. Kate’s eyesight is perfectly normal, and she sees her watch clearly. Kate’s hearing is perfectly normal, and she hears the passerby quite well. She has no special reason to believe that the passerby is inaccurate. She also has no special

reason to believe that her watch is inaccurate. Kate could gather further evidence that she is on Main Street (she could, for instance, find a map), but she doesn't do so, since, on the basis of what the passerby tells her, she already thinks that she is on Main Street.

Important (Passerby):

Kate needs to get to Main Street by noon: her life depends upon it. She comes to an intersection and asks a passerby the name of the street. "Main Street," the passerby says. Kate looks at her watch, and it reads 11:45 AM. Kate's eyesight is perfectly normal, and she sees her watch clearly. Kate's hearing is perfectly normal, and she hears the passerby quite well. She has no special reason to believe that the passerby is inaccurate. She also has no special reason to believe that her watch is inaccurate. Kate could gather further evidence that she is on Main Street (she could, for instance, find a map), but she doesn't do so, since, on the basis of what the passerby tells her, she already thinks that she is on Main Street.

Unimportant (Drunk):

Kate is ambling down the street, out on a walk for no particular reason and with no particular place to go. She comes to an intersection. A drunk is standing on the corner, looking a bit unsteady on his feet. Kate asks the drunk the name of the street. "Main Street," the drunk says. Kate looks at her watch, and it reads 11:45 AM. Kate's eyesight is perfectly normal, and she sees her watch clearly. Kate's hearing is perfectly normal, and she hears the drunk quite well. She has no special reason to believe that the drunk is inaccurate. She also has no special reason to believe that her watch is inaccurate. Kate could gather further evidence that she is on Main Street (she could, for instance, find a map), but she doesn't do so, since, on the basis of what the drunk tells her, she already thinks that she is on Main Street.

Important (Drunk):

Kate needs to get to Main Street by noon: her life depends upon it. She comes to an intersection. A drunk is standing on the corner, looking a bit unsteady on his feet. Kate asks the drunk the name of the street. "Main Street," the drunk says. Kate looks at her watch, and it reads 11:45 AM. Kate's eyesight is perfectly normal, and she sees her watch clearly. Kate's hearing is perfectly normal, and she hears the drunk quite well. She has no special reason to believe that the drunk is inaccurate. She also has no special reason to believe that her watch is inaccurate. Kate could gather further evidence that she is on Main Street (she could, for instance, find a map), but she doesn't do so, since, on the basis of what the drunk tells her, she already thinks that she is on Main Street.

Unimportant (Street Signs):

Kate is ambling down the street, out on a walk for no particular reason and with no particular place to go. At 11:45 AM, she comes to an intersection and looks up at a street sign that says "Main Street." She looks at her watch, and it reads 11:45 AM. The ambient lighting is perfectly normal. Kate's eyesight is perfectly normal, and

she sees the street sign clearly and sees her watch clearly. She has no special reason to believe that the street sign is inaccurate. She also has no special reason to believe that her watch is inaccurate. Kate could gather further evidence that she is on Main Street (she could, for instance, ask a passerby or find a map), but she doesn't do so, since, on the basis of the street sign, she already thinks that she is on Main Street.

Important (Street Signs):

Kate needs to get to Main Street by noon: her life depends upon it. At 11:45 AM, she comes to an intersection and looks up at a street sign that says "Main Street." She looks at her watch, and it reads 11:45 AM. The ambient lighting is perfectly normal. Kate's eyesight is perfectly normal, and she sees the street sign clearly and sees her watch clearly. She has no special reason to believe that the street sign is inaccurate. She also has no special reason to believe that her watch is inaccurate. Kate could gather further evidence that she is on Main Street (she could, for instance, ask a passerby or find a map), but she doesn't do so, since, on the basis of the street sign, she already thinks that she is on Main Street.

The following factors were available to participants in the general commitment test: how good Kate's hearing is; the quality of the street's asphalt; Kate's emotional state; the passerby's emotional state; how important it is that Kate be on Main Street at noon; the passerby's race; whether or not Kate speaks English; how often people have given Kate reliable information in the past; and the country in which Kate lives.