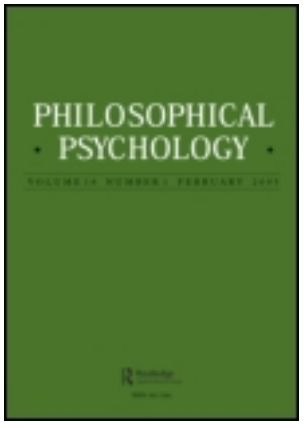


Philosophical Psychology
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/cphp20

# Tracking instability in our philosophical judgments: Is it intuitive?

Jennifer Wright

Published online: 29 Mar 2012.

To cite this article: Jennifer Wright (2013) Tracking instability in our philosophical judgments: Is it intuitive?, Philosophical Psychology, 26:4, 485-501, DOI: 10.1080/09515089.2012.672172

To link to this article: http://dx.doi.org/10.1080/09515089.2012.672172

*Philosophical Psychology*, 2013
Vol. 26, No. 4, 485–501, http://dx.doi.org/10.1080/09515089.2012.672172



# Tracking instability in our philosophical judgments: Is it intuitive?

Jennifer Wright

*Skepticism about the epistemic value of intuition in theoretical and philosophical inquiry fueled by the empirical discovery of irrational bias (e.g., the order effect) in people's judgments has recently been challenged by research suggesting that people can introspectively track intuitional instability. The two studies reported here build upon this, the first by demonstrating that people are able to introspectively track instability that was experimentally induced by introducing conflicting expert opinion about certain cases, and the second by demonstrating that it was the presence of instability—not merely the presence of conflicting information—that resulted in changes in the relevant attitudinal states (i.e., confidence and belief strength). The paper closes with the suggestion that perhaps the best explanation for these (and other) findings may be that intuitional instability is not actually "intuitional."*

*Keywords: Confidence; Intuitional Stability; Skepticism*



## 1. Introduction

Just as the scientific investigation of the natural world involves certain core methodological procedures, the heart of philosophical inquiry has always involved the consultation of one's rational intuitions. One way this methodology is employed is through the generation and discussion of concrete cases meant to elicit intuitions that either support or challenge particular philosophical claims and/or theories. The Gettier (1963) cases are a well-known example of this—many view the intuitions they elicited to have effectively undermined our centuries-old theory of knowledge.

Intuition's central place in philosophical methodology notwithstanding, there are many who have voiced skepticism about the reliance on intuitions in philosophical discourse, arguing that intuitions are epistemically suspect for a variety of theoretical

Jennifer Wright is an Assistant Professor at the College of Charleston.
Correspondence to: Jennifer Wright, College of Charleston, 57 Coming Street, Charleston, SC 29424, USA. Email: wrightjj1@cofc.edu

and/or empirical reasons (e.g., Alexander & Weinberg, 2007; Cummins, 1998; Gendler, 2007; Hintikka, 1999, 2001; Machery, Mallon, Nichols, & Stich, 2004; Nichols & Knobe, 2007; Nichols, Stich, & Weinberg, 2003; Nisbett, Peng, Choi, & Norenzayan, 2001; Redelmeier & Shafir, 1995; Weinberg, 2007; Weinberg, Nichols, & Stich, 2001; Williamson, 2004), and recently this skepticism has been reinforced by empirical evidence suggesting that people's intuitive judgments can be manipulated by information that is (at least seemingly) irrelevant. Specifically, Swain, Alexander, and Weinberg (2008)—and, more recently, Wright (2010) and Zamzow and Nichols (2010)—found people's intuitive judgments about standard philosophical concrete cases to be vulnerable to a sort of "order effect," their judgments about those cases varying significantly, depending upon which other cases they had been exposed to immediately prior to reading the cases in question.[1]

In response to empirical findings such as these, Weinberg (2007) and others have expressed a worry about the epistemic status of intuitions—at least, insofar as they are employed by philosophers as evidence for/against various philosophical claims—arguing that intuitions are epistemically "hopeless"; that, unlike other fundamental sources of information (e.g., perception), we have no good account of the processes involved and no good way of anticipating when they will lead us astray or what will cause them to do so. This leaves us currently unable to sufficiently "calibrate" our intuitive system, to protect against the various factors that might bias our intuitive judgments—which makes their frequent use in philosophical discourse problematic and worth treating with suspicion (or, at the very least, extreme caution).

There have been many philosophical responses to this skepticism (e.g., D. Sosa, 2006; E. Sosa, 2000, 2007, 2009). There have also been recent empirical findings that suggest that intuition may not be completely hopeless after all. Studies conducted by Wright (2010) and Zamzow and Nichols (2010), for example, provided preliminary evidence that not only are only *some* intuitive judgments unstable but also that methods for anticipating this instability do, in fact, exist. Specifically, these studies suggest that people have a reliable means by which to track intuitional instability, insofar as they possess introspective access to certain attitudinal states (e.g., the confidence and strength of belief in their judgments) that can serve as indicators of when our intuitive judgments are vulnerable to bias.

As Wright (2010) argued, when presented with cases that elicit only unclear/weak intuitions (or no intuitions at all) people's judgments become vulnerable to instability—i.e., vulnerable to the influence of seemingly irrelevant information, such as the case they had just previously read (e.g., whether or not someone "knows x" in a previous case should have no bearing on whether or not someone else "knows x" in a completely unrelated case). In such cases, people experience less confidence, and believe less strongly in, their resulting judgments. On the other hand, in the presence of clear/strong intuitions—such as those elicited by paradigmatic cases[2]—no instability is generated, and people's confidence/belief strength in their judgments is high.

Such findings provide important support against the skeptical undermining of philosophical methodology. But, as yet, these results are purely correlational—they

do not tell us whether the presence of this instability *results in* reduced confidence/belief strength. Perhaps the unclear cases used are just the sorts of cases people are likely to feel less confident about (and have weaker belief strength for) independently of the fact that they are vulnerable to instability. In order to further explore the relationship between instability and the relevant attitudinal states, we need to investigate whether confidence/belief strength will continue to track instability when it is generated where it has not been previously found—namely, in clear cases that had previously been shown to elicit clear/strong intuitions and to elicit high degrees of confidence/belief strength. Will an experimental induction of instability result in reductions in participants' confidence/belief strength for their judgments? The two studies discussed here were designed to experimentally answer this question.



## 2. Study 1

Study 1 explored the relationship between the stability of people's intuitive judgments and certain attitudinal states by experimentally generating instability using cases that have previously been shown to elicit clear/strong intuitions. It was important to employ such cases, because people have previously been shown to possess a high degree of confidence/belief strength in their judgments about them. Therefore, if instability in people's intuitive judgments about such cases could be successfully induced, we'd be able to measure any corresponding changes in their attitudinal states, to see if the former results in a reduction in the latter. The hypothesis was that this experimental induction of instability would indeed result in a corresponding reduction in participants' reported confidence/belief strength levels.

215 undergraduate college students (159 females; dominantly Caucasian) from the College of Charleston participated in this study. Participants were recruited through the Introduction to Psychological Science research pool and received research credit for their participation. Seven participants were eliminated from the study due to incomplete surveys and all analyses were conducted with the remaining 208 participants. Since participants' prior philosophical training had previously been found to not be predictive of differences in intuitional stability (Wright, 2010) and such training is difficult to obtain in sufficient quantities in college samples, this question was not asked here.

The goal of this study was to induce instability using cases that had previously been shown to elicit clear/strong intuitions. Four cases were chosen (see Appendix), two cases in epistemology and two in ethics. All four cases had been shown in previous research to elicit stable "yes" or "no" judgments—that is, people's intuitive judgments about these cases were not vulnerable to the order effect (Swain et al., 2008; Wright, 2010). The four cases used are as follows:

*Epistemological cases:*

**Clear yes** (*Perception*): Pat walks into her kitchen during the day when the lighting is good and there is nothing interfering with her vision. She sees a red apple sitting on the counter, where she had left it after buying it at the grocery store the

day before. As she leaves home, she tells her son, Joe, that there is a red apple sitting on the kitchen counter and to make sure to pack it with his lunch.

**Clear no** (*Coin-Flip*): Dave likes to play a game with flipping a coin. He sometimes gets a ''special feeling'' that the next flip will come out heads. When he gets this ''special feeling,'' he is right about half the time, and wrong about half the time. Just before the next flip, Dave gets that ''special feeling,'' and the feeling leads him to believe that the coin will land heads. He flips the coin, and it does land heads.

*Ethical cases:*

**Clear yes** (*Break-Promise*): Stan promises his grandfather that he will give him a ride to a free clinic at the hospital for his annual check-up at 12pm on Wednesday. Wednesday at 11:45am, on his way to his grandfather's house, Stan gets a call from his friend, who is on his way to a baseball game. He has an extra ticket, and invites Stan to join him. Stan decides to go with his friend to the game, even though he knows that doing so means that he will be breaking his promise to take his grandfather to the free clinic for his annual check-up.

**Clear no** (*Hide-Neighbors*): Hilda hides her Jewish neighbors in her basement during the Nazi occupation of France. A German soldier comes to her door one afternoon and asks her if she knows where her neighbors have gone. Hilda lies to the soldier, telling them no, she hasn't seen them recently, but she believes that they fled the country.

Since for all four cases, the order in which they were presented did not affect participants' judgments, instability needed to be induced in some other way. In previous instances of instability the presence of unclear/weak intuitions—and/or the lack of any intuition—made people's judgments relatively easy to manipulate. These cases, however, had been shown to elicit clear/strong intuitions, which meant that generating instability required providing information that would either interfere with, or cause people to override, those intuitions.

In order to do this, expert opinion information was introduced and manipulated—that information being either *consistent* or *inconsistent* with the dominant case (''yes'' or ''no'') intuition. Strictly speaking, whether or not a concrete case elicits a particular intuition (that is, whether or not the case directly strikes a person as an instance of knowledge or as morally wrong) should be impervious to other people's (dis)agreeing opinions, even when those other people are experts. Simply put, a case strikes *you* as it strikes you, regardless of how it strikes others, even if you ultimately decide to ignore/override that information. Thus, to the extent that people's judgments about these cases were manipulated by expert opinion, we have good reason to believe that they were forming those judgments on the basis of something other than their own intuitions.

While arguably a different way of inducing instability than manipulating the order of presentation of unclear cases (for which judgments are more easily manipulated due to the lack of any clear/strong intuitions to begin with), this approach is nonetheless similar insofar as people are forming judgments about cases on the basis of something *other than* clear/strong intuitions. In previous research, people's judgments were most likely influenced by something about the previously considered

case; here their judgments were being influenced by conflicting information about expert opinion. The expectation was that, to the extent that their judgments were so influenced, their attitudinal states would reflect that—a fact that should be introspectively trackable.[3]

While every participant received all four cases, they did so under one of four different conditions which varied along two dimensions: (1) whether they received the clear "yes" or clear "no" cases first; and (2) whether they were given *consistent* or *inconsistent* expert opinions about the cases. The only thing that did not vary between the conditions was the order of the epistemological versus ethical cases: participants were always presented with the two epistemological cases preceding the two ethical cases.[4] Thus, in conditions 1 & 2, participants received the "yes" cases before the "no" cases (condition 1: Epistemology Yes/Epistemology No/Ethical Yes/Ethical No), while in conditions 3 & 4, it was the opposite. In conditions 1 & 3, participants received inconsistent expert opinion, while in conditions 2 & 4, they received consistent expert opinion.

To give an example, consider the participants who were presented with the *Perception* case: In the *consistent* version, participants read the following:

> When speaking with her son, did Pat know that there was a red apple on the counter in the kitchen? We gave this case to 10 professional epistemologists and linguists and they were divided—6 out of the 10 stated that "yes, Pat knew."

Then participants were asked:

> What do you think? When speaking with her son, did Pat know that there was a red apple on the counter in the kitchen?"

to which they could answer "yes" or "no."

In the *inconsistent* version, participants read the following:

> When speaking with her son, did Pat know that there was a red apple on the counter in the kitchen? We gave this case to 10 professional epistemologists and linguists and they were divided—7 out of the 10 stated that "no, Pat did not know."

Then participants were asked:

> What do you think? When speaking with her son, did Pat know that there was a red apple on the counter in the kitchen?

to which they could answer "yes" or "no."

After each case, participants were then asked how *confident* they were in their answer (1 = Not at all confident, to 7 = Very confident) and how *strongly they believed* their answer (1 = Not at all strongly, to 7 = Very strongly).[5]

### *2.1. Results*

Of the four cases participants considered, the experimental manipulation induced significant instability in three: *Perception, Coin-Flip,* and *Hide-Neighbors*. Participants attributed knowledge to Pat in *Perception* more frequently when given a consistent expert opinion (89%) than when given an inconsistent expert opinion (77%),

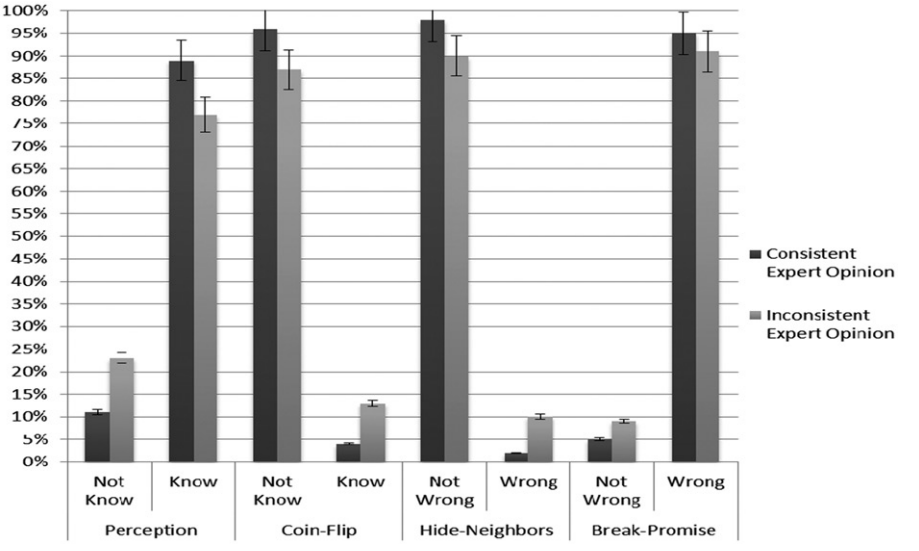


**Figure 1** Participants' Judgments About Cases when Provided with Consistent Versus Inconsistent Expert Opinion Information; SE bars included.

$X^2(1, N = 206) = 4.7$, $p = 0.03$, and failed to attribute knowledge to Dave in *Coin-Flip* more frequently when given a consistent expert opinion (96%) then when given an inconsistent expert opinion (87%), $X^2(1, N = 208) = 5.9$, $p = 0.015$. In addition, participants judged Hilda's action in *Hide-Neighbors* as not wrong more frequently when given a consistent expert opinion (98%) than when given an inconsistent expert opinion (90%), $X^2(1, N = 208) = 6.3$, $p = 0.012$. They did not, however, judge Stan's action in *Break-Promise* as wrong more frequently when given a consistent expert opinion (95%) than when given an inconsistent expert opinion (91%), $X^2(1, N = 208) = 1.6$, $p = 0.21$, though the trend was moving in the right direction (i.e., more participants judging Stan's action as wrong following consistent than inconsistent expert opinion; Figure 1).[6]

It was not particularly surprising to see somewhat less movement in the ethical cases than the epistemological cases. Previous research (Skitka, Bauman, & Lytle, 2009) has shown that people are less willing to defer to conflicting authority opinion when it comes to their moral (as opposed to non-moral) judgments. And it makes sense that it would be easier to influence people's judgments about Hilda's action than it would be Stan's—while Hilda's action took place within a larger context of political conflict and could be rationally argued to be wrong for a variety of reasons (such as the potential negative ramifications for herself and others), Stan's action was purely selfish and inconsiderate of his grandfather's wellbeing and, thus, would be hard to argue was *not* wrong.

When compared to the judgments of participants who were not given any expert opinion information (Wright, 2010), we see interesting differences between participants' judgments when they were given no expert information and when

**Table 1** Difference in Confidence and Belief Strength Between Study 1, 'Expert Opinion', and Wright (2010), 'No Expert Opinion'

| Case | Introspected states | Condition | *N* | Mean | SE | *T* | df | *p* value |
|---|---|---|---|---|---|---|---|---|
| Perception | Confidence | Expert opinion | 208 | 5.98 | 0.08 | | | |
| | Confidence | No expert opinion | 177 | 6.31 | 0.07 | −2.9 | 383 | 0.004 |
| | Belief strength | Expert opinion | 207 | 6.00 | 0.08 | | | |
| | Belief strength | No expert opinion | 177 | 6.28 | 0.07 | −2.5 | 382 | 0.011 |
| Coin-Flip | Confidence | Expert opinion | 208 | 6.12 | 0.08 | | | |
| | Confidence | No expert opinion | 177 | 6.72 | 0.05 | −6.1 | 383 | 0.000 |
| | Belief strength | Expert opinion | 206 | 6.14 | 0.08 | | | |
| | Belief strength | No expert opinion | 176 | 6.73 | 0.05 | −6.1 | 380 | 0.000 |
| Hide-Neighbors | Confidence | Expert opinion | 208 | 6.29 | 0.08 | | | |
| | Confidence | No expert opinion | 176 | 6.69 | 0.06 | −4 | 382 | 0.000 |
| | Belief strength | Expert opinion | 208 | 6.34 | 0.07 | | | |
| | Belief strength | No expert opinion | 177 | 6.67 | 0.06 | −3.5 | 383 | 0.001 |

they were given either consistent or inconsistent expert opinions. For example, 84% of participants attributed knowledge in *Perception* when given no information, which was less than when they were given a consistent expert opinion and more than when they were given inconsistent expert opinion, $X^2(2, N = 382) = 4.9$, $p = 0.08$ and in *Coin-Flip,* participants failed to attribute knowledge more frequently when they had no information (97%) or had been given a consistent expert opinion than when they were given an inconsistent expert opinion, $X^2(2, N = 383) = 13.0$, $p = 0.001$. Similarly, in *Hide-Neighbors,* they judged Hilda's action to not be wrong more frequently when they had no information (97%)[7] or had been given a consistent expert opinion than when they were give an inconsistent expert opinion, $X^2(2, N = 385) = 10.1$, $p = 0.006$.[8]

Of course, the most important question is whether this experimental induction of instability in cases that had previously elicited clear/strong intuitions led to a corresponding reduction in participants' confidence and/or belief strength in their judgments. If these introspectively accessible attitudinal states did indeed track instability, then we should see a downward shift in reported confidence/belief strength from these same cases when they had displayed stability (as they had in Wright, 2010). And, indeed, this is what was found. Participants from the Wright (2010) studies, who had received the *Perception, Coin-Flip,* and *Hide-Neighbors* cases without expert opinion information, showed a higher level of confidence and belief strength in their judgments than participants who had received either consistent or inconsistent expert opinion information (Table 1; see also note 8).

It is interesting to note that participants' confidence was lower in *both* experimental conditions, whether they received the consistent or inconsistent expert opinion information. In fact, there was no significant difference between the confidence/belief strength reported between the consistent and inconsistent conditions, *t*s(206–216) = 0.084–1.63, *ns*. This suggests that while the introduction of consistent versus inconsistent expert opinions generated instability in participants'

judgments (participants gave the dominant answer more frequently in the no information and consistent conditions than in the inconsistent condition), it was the induction of instability itself—via the introduction of external information that influenced their judgments—that resulted in a reduction in participants' confidence and belief strength. Whether this information *supported* or *undermined* their initial intuitions did not matter.

### *2.2. Discussion*

Study 1 showed that the experimental induction of instability in cases that had previously elicited stable judgments resulted in a corresponding reduction in participants' confidence/belief strength in their judgments. These findings provide further support for the view that people's introspective access to certain attitudinal states serves as a reliable indicator of the instability of their judgments. In addition, it suggests that this instability can be generated in different ways for different reasons: e.g., because people lack clear/strong intuitions (having only weak/unclear intuitions or no intuitions at all) about a particular case—and, therefore, rely on other information to form their judgment; or because other information interferes with their clear/strong intuitions, causing them to question/ignore/override them. Importantly, however the instability is generated, there are certain attitudinal states that reliably reflect its presence. In short, people experience less confidence/belief strength for judgments that they form on the basis of information that interferes with, or takes the place of, clear/strong intuitions.

There is, however, an important worry about study 1's design, which is that it may have induced instability in previously stable cases by introducing uncertainty about the correct answer. While the consistent case differed from the inconsistent case in terms of whether the majority of experts came down on the same or the opposing side, they nonetheless both introduced uncertainty insofar as the experts were represented as *disagreeing* with one another (e.g., with *Perception,* the *consistent* case had 6 out of 10 experts saying "yes, Pat knew" and the *inconsistent* case had 7 out of 10 experts saying "no, Pat didn't know").

This is a worry for two reasons. First, it is always possible that it was the presence of this uncertainty that generated not only the instability (which was the intention), but also the reduction in participants' confidence/belief strength. After all, it is easy to lose confidence in your judgment when you have just been made aware of the fact that there are some experts out there that disagree with you and/or with each other. Second, if this is the case, then it seems reasonable to expect that the presence of widespread expert *agreement* about a case would boost participants' confidence/belief strength in their intuitive judgments, especially if those judgments are consistent with said expert opinion (Koriat, 2008).

This would be problematic for the view being argued for here (and in Wright, 2010) in the following way: it would mean that instability could be generated with no corresponding drop in participants' confidence/belief strength. If instability in people's judgments was generated through the use of expert opinion (as in study 1),

only now with expert opinion that was either consistently *for* or *against* a particular judgment—which would likely result in increased confidence/belief strength in the participants—then this would mean that people's introspectively accessible attitudinal states would no longer serve as indicators of instability; their ability to track instability having been effectively (and quite easily) derailed.

However, if it is the instability itself that is resulting in the reduction in confidence/belief strength, then we should expect to see a corresponding decrease (*not* an increase) in the relevant attitudinal states. Study 2 was designed to test this possibility.



## 3. Study 2

109 undergraduate college students (83 females; dominantly Caucasian) from the College of Charleston participated in this study. Participants were recruited through the Introduction to Psychological Science research pool and received research credit for their participation.

This time participants were presented with two different sets of cases; six cases in total. One set involved three cases in epistemology and the other, three ethical cases. As in study 1, four of the six cases were ''clear'' cases (i.e., cases that had been previously shown to elicit stable ''yes'' or ''no'' judgments), while the other two cases were ''unclear'' cases (i.e., cases that had been previously shown to be vulnerable to the order effect; Swain, et al., 2008; Wright, 2010). The cases used were as follows:

*Epistemological cases:*

**Clear yes** (*Testimony*): Karen is a distinguished professor of chemistry. This morning, she read an article in a leading scientific journal that mixing two common floor disinfectants, Cleano Plus and Washaway, will create a poisonous gas that is deadly to humans. In fact, the article is correct: mixing the two products does create a poisonous gas. At noon, Karen sees a janitor mixing Cleano Plus and Washaway and yells to him, ''get away! Mixing those two products creates a poisonous gas!''

**Clear no** (*Coin-Flip*): Dave likes to play a game with flipping a coin. He sometimes gets a ''special feeling'' that the next flip will come out heads. When he gets this ''special feeling,'' he is right about half the time, and wrong about half the time. Just before the next flip, Dave gets that ''special feeling,'' and the feeling leads him to believe that the coin will land heads. He flips the coin, and it does land heads.

**Unclear** (*True-Temp*): Suppose Charles undergoes brain surgery by an experimental surgeon who invents a small device which is both a very accurate thermometer and a computational device capable of generating thoughts. The device, called a tempucomp, is implanted in Charles' head so that the very tip of the device, no larger than the head of a pin, sits unnoticed on his scalp and acts as a sensor to transmit information about the temperature to the computational system of his brain. This device, in turn, sends a message to his brain causing him to think of the temperature recorded by the external sensor. Assume that the tempucomp is very reliable, and so his thoughts are correct temperature thoughts. All told, this is a reliable belief-forming process. Charles has no idea that the tempucomp has been

inserted in his brain, is only slightly puzzled about why he thinks so obsessively about the temperature, but never checks a thermometer to determine whether these thoughts about the temperature are correct. He accepts them unreflectively, another effect of the tempucomp. Thus, at a particular moment in time he thinks and accepts that the temperature is 71 degrees—and it is, in fact, 71 degrees.

*Ethical cases:*

**Clear yes** (*Sell-iPod*): Laura and Suzy are roommates. Laura asks Suzy if she has seen her new iPod, which she had worked an extra job over the summer to be able to afford. Suzy did recently see it under a pile of papers on the bookshelf. But Suzy lies to Laura, telling her that she hasn't seen it. She thinks that if Laura doesn't find it on her own in a day or two, she can take it down to the pawn shop and get $100 for it, which would provide her with beer money for the week.

**Clear no** (*Break-Promise*): Fred promises his girlfriend that he will meet her for lunch at 12pm on Wednesday at their favorite café. Wednesday at 11:45am, on his way to the café, Fred runs into his grandfather, who is out for a stroll. They exchange hellos, and then suddenly Fred's grandfather clutches his chest and falls to the ground unconscious. An ambulance arrives minutes later to take Fred's grandfather to the hospital. Fred accompanies his grandfather to the hospital, even though he knows that doing so means that he will be breaking his promise to have lunch with his girlfriend.

**Unclear** (*Hide-Bombers*): Martha hides her Jewish neighbors in her basement during the Nazi occupation of France. A German soldier comes to her door one afternoon and asks her if she knows where her neighbors have gone. Martha knows that her neighbors are wanted by the Germans for bombing a German-only schoolyard and killing several children, injuring others. Martha lies to the soldier, telling them no, she hasn't seen them recently, but she believes that they fled the country.

Participants were assigned to one of two conditions. The three epistemological cases always preceded the three ethical cases and the unclear cases were always preceded by a clear case, but whether that case was a clear "yes" or "no" case was counterbalanced (condition 1: Epistemology Yes-Unclear-No/Ethics Yes-Unclear-No; condition 2: Epistemology No-Unclear-Yes/Ethics No-Unclear-Yes). Each version presented participants with a case and then provided them with expert opinion about the case. The expert opinion was always in agreement and consistent with the clear "yes" and "no" cases. For instance, in *Testimony,* participants were told that:

> We asked over 100 professional epistemologists and linguists and they dominantly agreed that "yes, Karen knew."

In *Coin-Flip,* participants were told that:

> We asked over 100 professional epistemologists and linguists and they dominantly agreed that "no, Dave did not know."

Being presented with one of the clear cases first helped establish the "believability" of the expert opinion being provided, so that when participants were next presented with an unclear case they would feel more inclined to take it seriously. Since what

varied between versions was the expert opinion that was given for the unclear cases, this helped to generate instability.

In condition 1, both unclear cases (*True-Temp* and *Hide-Bombers*) were presented with a consistent "yes" expert opinion; in condition 2, the unclear cases were presented with a consistent "no" expert opinion.

As an example, consider the participants who were presented with the True-Temp case. In the "dominant yes" version participants read the following:

> Did Charles know that the temperature is 71 degrees? We asked over 100 professional epistemologists and linguists and they dominantly agreed that "yes, Charles knew."

Then participants were asked:

> What do you think? Did Charles know that the temperature is 71 degrees?

to which they could answer "yes" or "no."

In the "dominant no" version, participants read the following:

> Did Charles know that the temperature is 71 degrees? We asked over 100 professional epistemologists and linguists and they dominantly agreed that "no, Charles did not know."

Then participants were asked:

> What do you think? Did Charles know that the temperature is 71 degrees?'

to which they could answer "yes" or "no."

After each case, participants were then asked how *confident* they were in their answer (1 = Not at all confident, to 7 = Very confident) and how *strongly they believed* their answer (1 = Not at all strongly, to 7 = Very strongly).

### *3.1. Results*

As was expected, judgments for all four clear "yes" and "no" cases remained stable across both conditions—comparing participants' judgments for each case both when presented first (before an unclear case) and when presented third revealed no differences, $X^2(1, X = 108–109)$, s = 0.17–2.5, *p*s = *ns*. In addition, both of the unclear cases demonstrated the anticipated instability. In *True-Temp,* participants attributed knowledge to Charles more frequently when the expert opinion was a dominant "yes" (63%) than when it was a dominant "no" (35%), $X^2(1, X = 109) = 8.9$, $p = 0.003$. In *Hide-Bombers,* participants judged Hilda's action to be wrong more frequently when the expert opinion was a dominant "yes" (50%) than when it was dominant "no" (18%), $X^2(1, X = 109) = 11.7$, $p < 0.001$ (Figure 2).

But did being exposed to consistent expert opinion increase participants' confidence/belief strength in their answers? Importantly, it did *not*. Participants' reported confidence and belief strength remained significantly lower for both unclear cases (confidence *M*s = 5.7 and 5.9, *SE*s = 0.13, belief strength *M*s = 5.8 and 6.0, *SE*s = 0.12–0.13) than for any of the clear cases (confidence *M*s = 6.4–6.7, *SE*s = 0.06–

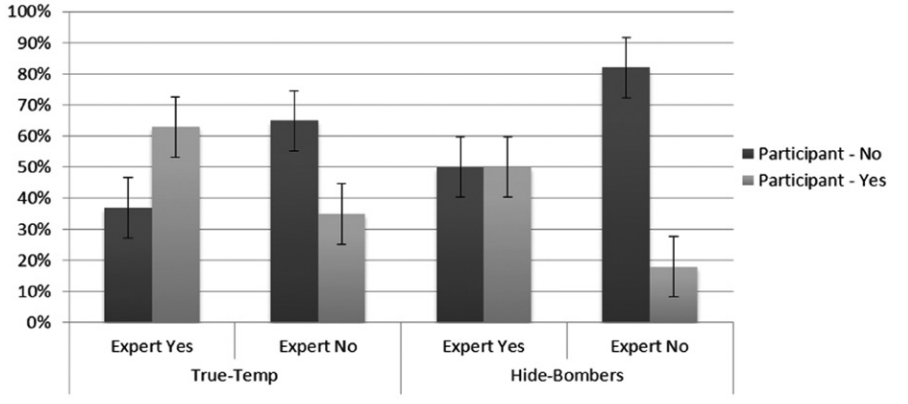


**Figure 2** Study 2: Participants' Judgments About Cases when Provided with Dominant "Yes" Versus "No" Expert Opinion Information; SE bars included.



0.11, belief strength $Ms = 6.3–6.7$, $SEs = 0.06–0.10$), $ts$ $(106–108) = 4.0–5.8$, $ps < 0.001$. In addition, there were no significant differences in either confidence or belief strength between the clear "yes" and "no" cases, $ts(106–107) = 0.38–0.71$, *ns* (Figure 3).

Of course, these analyses were collapsed across those participants whose judgments were consistent with the dominant expert opinion and those whose were not. Maybe the reason why participants' confidence and belief strength for the unclear cases was reduced is because the mean includes the attitudinal states reported for both judgments. To examine this, the confidence and belief strength for participants whose judgments in the unclear cases were consistent with the dominant expert opinion were compared to those whose judgments were inconsistent.

For *True-Temp,* participants whose judgments were consistent with the dominant expert opinion did express more confidence than those whose judgments were inconsistent, both when those judgments were affirmative ("yes, Charles knows"), $t(58) = 3.4$, $p = 0.001$, and when they were negative ("no, Charles doesn't know"), $t(47) = 2.2$, $p = 0.035$. Interestingly, though, participants' belief strength did not display this same variability—it was higher for participants' consistent judgments when those judgments were affirmative, $t(58) = 2.5$, $p = 0.016$, but not negative, $t(47) = 1.5$, *ns*. In addition, in *Hide-Bombers* there was no difference in either confidence or belief strength between participants whose judgments were consistent and those whose were inconsistent with the dominant expert opinion, whether they were affirmative or negative, $ts(47–58) = 0.56–1.3$, *ns*. So, while there was a difference between the attitudinal states reported by participants whose judgments were consistent with the presented expert opinion and those whose judgments were inconsistent with the presented expert opinion, it was limited to one attitudinal state for one of the unclear cases, so this clearly cannot fully account for the reduction in confidence and belief strength found in the unclear cases.

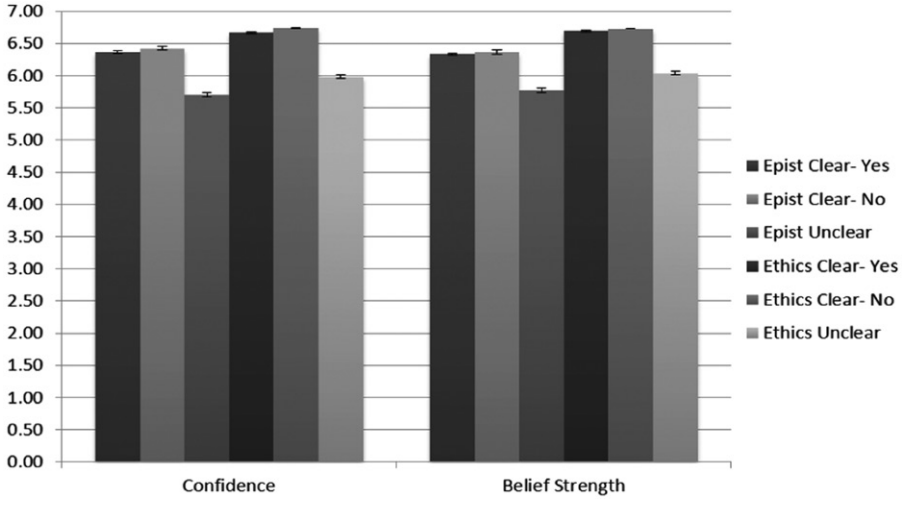


**Figure 3** Study 2: Confidence and Belief Strength for Clear (Yes/No) Versus Unclear Cases; SE bars included.

To reinforce this conclusion, comparisons were also made between participants' attitudinal states in the unclear cases, when their judgments were consistent with the presented expert opinion, and their attitudinal states in the clear "yes" and "no" cases. If what was driving the reduction in confidence and belief strength judgments in the unclear cases was simply a greater percentage of participants making judgments that were inconsistent with the expert opinion in the unclear cases than the clear cases, then we should expect participants' confidence/belief strength to be more or less identical between the unclear and clear cases when we compare *only* those participants whose judgments had been consistent with the presented expert opinion. Yet, the only place that where this was actually found was in the comparison between *True-Temp* and the clear "yes" and "no" cases when the dominant expert opinion in the unclear case was affirmative, $ts(36–37) = 0.10–1.1$, *ns*. Both when the dominant expert opinion was negative in *True-Temp,* and for both (affirmative/negative) conditions in *Hide-Bombers,* the attitudinal states that participants reported for the unclear were significantly lower than those reported for the clear cases, $ts(29–39) = 2.0–4.4$, $ps < 0.05$, even though everyone's judgments were consistent with the expert opinion presented.

### *3.2. Discussion*

Contrary to the worry raised by study 1, the induction of instability with expert agreement did *not* result in a decoupling of instability and the associated decrease in confidence and belief strength. Instead, once again, it was found that participants' attitudinal states to be appropriately responsive to the presence of instability, suggesting that it was the presence of the instability (however induced) and *not*

merely the presence of perceived (dis)agreement that generated participants' decreased confidence and belief strength.

## 4. General Discussion

Recently, philosophers (armed with empirical data) have suggested that philosophers' reliance on intuitions is problematic because they are vulnerable to various forms of bias—and worse, that there is no way to protect against this bias (Alexander & Weinberg, 2007; Swain, et al., 2008; Weinberg, 2007). The studies here, taken together, suggest that that people *do* have a way to track at least one important form of bias—namely, when their judgments are being formed on the basis of clear/strong intuitions and when they are not. These and other studies (e.g., Wright, 2010; Zamzow & Nichols, 2010) provide evidence that certain introspectible attitudinal states, such as confidence and belief strength, can serve as reliable indicators of instability, suggesting that such states could potentially serve as useful tools with which to protect our intuitive philosophical judgments from external manipulation.

Again, it is important to be clear that what we've been referring to as 'intuitional instability' may not be, strictly speaking, *intuitional*. That is, one possible explanation for these—and other—findings is that the instability generated in people's judgments about concrete cases is the result of their judgments being formed on the basis of something *other than* clear/strong intuitions, either because the intuitions they have are unclear/weak or they have no intuitions at all (such as in the unclear cases), or because their clear/strong intuitions are somehow being interfered with or overridden, as in the clear cases for which conflicting expert opinion was provided. In these cases, it may be that people are forming judgments on the basis of other forms of information (e.g., details of previous cases, presence of expert opinion), making those judgments more easily influenced.[9]

This possibility has important implications for the debate about the status of intuitions in philosophical methodology (Alexander & Weinberg, 2007; Cummins, 1998; Hintikka, 1999, 2001; Machery, Mallon, Nichols, & Stich, 2004; Nichols, Stich, & Weinberg, 2003; Weinberg, 2007; Weinberg, Nichols, & Stich, 2001; Williamson, 2004). If intuitions themselves turn out not to be the source of the problem—insofar as when people have clear/strong intuitions that are not somehow being interfered with/overridden, then they are less vulnerable to these sorts of biases—then this changes the nature of the debate. The target for critics of philosophical practices becomes not *intuition* per se (at least, not those that are clear/strong), but those judgment formation processes that, at various points, take the place of or interfere with our intuitions. Given that there is already a healthy body of literature out there on such processes and the cognitive biases they are vulnerable to (as well as what can be done to protect against them), those interested in protecting/improving philosophical methodology are in good hands.

## Notes

[1] Though here I've glossed over a variety of different types of criticisms of intuitions, it is important to be clear that the cultural and gender differences found in people's intuitions is a very different sort of problem than the problem of cognitive bias (such as the order effect)—and the studies discussed here are intended to speak only to the latter. The latter involves our intuitive judgments being unduly influenced by information we are privy to at the moment our judgments are formed, while the former involves the ways in which the socio-cultural norms we grow up within (and which we arguably internalize) shape our understanding and use of certain concepts—and perhaps even the concepts themselves. Though these differences may be at the heart of interesting cultural differences in our philosophical theories, we do not typically regard them as a form of "bias" (at least not in the way we regard the latter to be). Exploring the roots (and the scope) of these cross-cultural differences is nonetheless a very important philosophical endeavor.

[2] By "paradigmatic" cases, I mean cases generally recognized to be uncontroversial instances of particular concepts, the sorts of cases a person would use to illustrate a concept. For example, *Coin-Flip Dave* is standardly recognized as an uncontroversial example of *lacking* knowledge (indeed, in was introduced in Swain, et al., 2008, as a comprehension measure, to ensure people were employing the concept KNOW correctly). Likewise, to say "I know there is an apple on the table" when you see an apple on the table in front of you, as in the *Perception* case, is an uncontroversial example of *having* knowledge (or knowing).

[3] It is important at this point to recognize that judgments formed on the basis of something other than one's intuition are not *intuitive* judgments (rather, they're inferential judgments), in which case the instability originally identified by Swain et al. (2008) would not be, strictly speaking, "intuitional." This is important for at least two reasons: (1) because it becomes less clear that the bias philosophical intuitions are vulnerable to is *irrational*—while the order in which cases are presented certainly seems irrelevant to how a case strikes you, relying on information from previous cases to form judgments about a case for which you do not have a clear/strong intuition about does not seem particularly irrational, nor would taking other people's (especially experts) opinions into account; and (2) because it becomes less clear that it is *philosophical intuitions* that are themselves vulnerable to bias, as opposed to people's judgment formation processes more generally, which is already a well-recognized and well-studied fact that should not trouble philosophers—or philosophy—any more than it troubles anyone else.

[4] The order of the epistemological and ethical cases was not counterbalanced because to do so would have doubled the number of participants required for the study and it was not necessary to test the proposed hypothesis.

[5] These two measures were used both to maintain consistency with previous research (Wright, 2010) and because though strongly correlated with one another (0.87–0.91), they are conceptually distinct (see Wright, 2010 for discussion).

[6] Participants' confidence/belief strength was higher in the *Break-Promise* case, which did not display significant instability (though it did display a trend in that direction), than in either of the *Perception* or *Coin-Flip* cases: $ts(207) = 2.8–4.9$, $ps < 0.005$, though not in the *Hide-Neighbors* case, $ts(207) = 0.4–0.8$, *ns*.

[7] The *Hide-Neighbors* data was collected as a part of an unpublished pilot study following Wright (2010).

[8] The *Break-Promise* case used in study 1 differed in several ways from the case used in Wright (2010), so direct comparisons between the two were not possible.

[9] Of course, this raises the question of how we can distinguish between when people's judgments are being formed on the basis of an intuition (i.e., when people report their intuitions—how something seems or strikes them—as their considered judgments) and when they are not. When people are asked to report their judgments (e.g., "does Pat know?

Yes or no''), we have no direct way of assessing the degree to which particular intuitions are present and/or involved in the judgments provided—and it is not clear what such an assessment would look like. In some unpublished data I've collected, I asked people to report how a case seemed to them/ struck them (including allowing them to say that the case struck them as both yes/no *or* as neither yes/no). Interestingly, the results looked very similar—for example, when given *True-Temp,* most people reported that it seemed to them that either Charles *knew* or *did not know* (instead of both or neither), and those judgments were still unstable, varying as a function of which case they'd seen first. Is this evidence for *intuitional* instability? It is hard to say. Clearly, more work needs to be done to clarify if and when intuitions (especially clear/strong ones) are involved in people's concrete case judgments.

## References

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

Cummins, R. (1998). Reflection on reflective equilibrium. In M. DePaul & W. Ramsey (Eds.), *Rethinking intuition: The psychology of intuition and its role in philosophical inquiry* (pp. 113–128). Lanham, MD: Rowman and Littlefield.

Gendler, T. (2007). Philosophical thought experiments, intuitions, and cognitive equilibrium. *Midwest Studies in Philosophy*, *31*, 68–89.

Gettier, E.L. (1963). Is justified true belief knowledge? *Analysis*, *23*, 121–123.

Hintikka, J. (1999). The emperor's new intuitions. *Journal of Philosophy*, *96*, 127–147.

Hintikka, J. (2001). Intuitionistic logic as epistemic logic. *Synthese*, *127*, 7–19.

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

Machery, E., Mallon, R., Nichols, S., & Stich, S. (2004). Semantics, cross-cultural style. *Cognition*, *923*, 1–12.

Nichols, S., & Knobe, J. (2007). Moral responsibility and determinism: The cognitive science of folk intuitions. *Nous*, *41*, 663–685.

Nichols, S., Stich, S., & Weinberg, J. (2003). Metaskepticism: Meditations in ethno-epistemology. In S. Luper (Ed.), *The skeptics* (pp. 227–247). Aldershot, UK: Ashgate Publishing.

Nisbett, R.E., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: Holistic versus analytic cognition. *Psychological Review*, *1082*, 291–310.

Redelmeier, D., & Shafir, E. (1995). Medical decision making in situations that offer multiple alternatives. *Journal of the American Medical Association*, *273*, 302–305.

Skitka, L.J., Bauman, C.W., & Lytle, B.L. (2009). Limits on legitimacy: Moral and religious convictions as constraints on deference to authority. *Journal of Personality and Social Psychology*, *97*, 567–578.

Sosa, D. (2006). Scepticism about intuition. *Philosophy*, *81*, 633–647.

Sosa, E. (2000). Replies. *Nous*, *10*, 38–42.

Sosa, E. (2007). Intuitions: Their nature and epistemic efficacy. *Grazer Philosophische Studien*, *74*, 51–67.

Sosa, E. (2009). A defense of the use of intuitions in philosophy. In M. Bishop & D. Murphy (Eds.), *Stich and his critics* (pp. 101–112). Oxford: Blackwell.

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

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

Weinberg, J., Nichols, S., & Stich, S. (2001). Normativity and epistemic intuitions. *Philosophical Topics*, *291*, 429–460.

Williamson, T. (2004). Philosophical ''intuitions'' and skepticism about judgment. *Dialectica*, *58*, 109–153.

Wright, J.C. (2010). Intuitional stability: The clear, the strong, and the paradigmatic. *Cognition*, *115*, 491–503.

Zamzow, J., & Nichols, S. (2010). Variations in ethical intuitions. *Philosophical Issues*, *19*, 368–388.