

Unstable Intuitions and Need for Cognition: How Being Thoughtful Sometimes Just Means Being Wrong in a Different Way

Jonathan M. Weinberg, Joshua Alexander, and Chad Gonnerman
Indiana University, Bloomington

Abstract

Some recent work by experimental philosophers has purported to raise the challenge that the kinds of hypothetical case intuitions so often used by analytic philosophers are, in fact, *unstable* and ought not to be trusted. One line of response to that work is that the experimentalists' subjects have only reports quick, unthoughtful judgments, but philosophers' intuitions are the product of greater thought and reflection, and thus can be expected to avoid that instability. This line of response is tested empirically, and found wanting – and in an unexpected way.

Introduction

The practice of treating intuitions as evidence has a long history in philosophy, arguably going as far back as Socrates.¹ Philosophers engaged in this practice consult intuitions generated in response to hypothetical cases as evidence for the truth or falsity of a wide variety of philosophical claims. In contemporary philosophy, hypothetical case intuitions are appealed to in order to support, or contest, philosophical claims about the nature of knowledge, justification, meaning, moral responsibility, and morally right action—to name just a few. In fact, the practice of treating hypothetical case intuitions as evidence is so widespread in philosophy that at least one philosopher has called it “our standard justificatory procedure”.²

¹ For examples of this practice from the history of philosophy, see M. DePaul and W. Ramsey, *Rethinking Intuition*, (Rowman and Littlefield, 1998); J. Pust, *Intuitions as Evidence*, (Garland Publishing, 2000); A. Goldman, “Philosophical Intuitions: Their Target, Their Source, and Their Epistemic Status”, *Grazer Philosophische Studien*, (forthcoming); and J. Nagel, “Epistemic Intuitions”, *Philosophy Compass*, (forthcoming).

² See G. Bealer, “The Incoherence of Empiricism”, *Proceedings of the Aristotelian Society*, 66, (1992): pp. 99-138; G. Bealer, “A Priori Knowledge and the Scope of Philosophy”, *Philosophical Studies*, Vol. 66, (1996): pp. 121-142; and G. Bealer, “Intuition and the Autonomy of Philosophy”, in *Rethinking Intuition*, eds M. DePaul and W. Ramsey, (Rowman and Littlefield, 1998): pp. 201-240. Of course, treating intuitions as evidence is not the only way of conceiving the role that intuitions can, or should, play in philosophical practice. For example, Hilary Kornblith has recently argued that intuitions should be treated more as hypotheses, or starting points, that inform the direction our early empirical investigations. See, in particular, H. Kornblith, “Naturalism and Intuitions”, *Grazer Philosophische Studien* (forthcoming) and H. Kornblith, *Knowledge and Its Place in Nature*, (Oxford University Press, 2002).

This practice of treating hypothetical case intuitions as evidence in philosophy has recently been challenged by a growing movement in philosophy—*experimental philosophy*.³ A series of recent empirical studies have been conducted by experimental philosophers on hypothetical case intuitions.⁴ These studies suggest that some particularly prominent, and commonly appealed to, hypothetical case intuitions have epistemically undesirable properties. In particular, these studies suggest that some hypothetical case intuitions are sensitive to philosophically non-salient features such as who is considering the hypothetical case, the presence or absence of affective content, and the context in which the hypothetical case is being considered.⁵ We will use the term “instability” to pick out this epistemically undesirable characteristic. Even though only a few hypothetical case intuitions have been examined, we can neither explain what it is about any of these cases that makes them unstable nor predict, of any other hypothetical case intuition to which we might like to appeal, whether or not that intuition will also be unstable. These findings thus present a challenge to the well-functioning of the philosophical practice of appealing to such intuitions. We’ll call this the “Experimentalists’ Challenge”.

Considering the implications this challenge has for how we go about the practice of doing philosophy, it is perhaps unsurprising that the Experimentalists’ Challenge has not been universally well-received. Philosophers intent on continuing the unrestricted practice of appealing to hypothetical case intuitions as evidence have attempted to mount a number of defenses.⁶ In this paper, we

³ For further discussion, see J. Alexander and J. Weinberg, “Analytic Epistemology and Experimental Philosophy”, *Philosophy Compass*, Vol. 2, (2007): pp. 56-80; J. Weinberg, “How to Challenge Intuitions Empirically Without Risking Skepticism”, *Midwestern Studies in Philosophy*, Vol. 31, (2007): pp. 318-343; and T. Nadelhoffer and E. Nahmias, “The Past and Future of Experimental Philosophy”, *Philosophical Explorations*, Vol. 10(2), (2007): pp. 123-149.

⁴ J. Weinberg, S. Nichols, and S. Stich, “Normativity and Epistemic Intuitions”, *Philosophical Topics*, 29, (2001): pp. 429-60; E. Machery, R. Mallon, S. Nichols and S. Stich, “Semantics, Cross-Cultural Style.” *Cognition*, Vol. 92, (2004): B1-B12; S. Nichols and J. Knobe, “Moral Responsibility and Determinism: The Cognitive Science of Folk Intuitions”, *Nous*, (forthcoming); and S. Swain, J. Alexander, and J. Weinberg, “The Instability of Philosophical Intuitions: Running Hot and Cold on Truetemp”, *Philosophy and Phenomenological Research*, (forthcoming).

⁵ And in ways for which philosophical forms of contextualism may not be appropriate; see below.

⁶ See, for example, A. Goldman, “Replies to the Contributors”, *Philosophical Topics*, 29, 2001, pp. 461-511; F. Jackson, “Responses”, *Philosophy and Phenomenological Research*, 62 (3), 2001, pp. 653-664; A. Kauppinen, “The Rise and Fall of Experimental Philosophy”, *Philosophical Explorations*, Vol. 10(2), (2007): pp. 95-118; S. Matthew Liao, “A Defense of Intuitions”, *Philosophical Studies*, (forthcoming); E. Sosa, “A Defense of the Use of Intuitions in Philosophy”, in *Stich and His Critics* eds M. Bishop and D. Murphy, (Blackwell Publishers, 2005); E. Sosa, “Experimental Philosophy and Philosophical Intuition”, *Philosophical Studies*, Vol. 132(1), (2007): pp. 99-107; and T. Williamson, “Philosophical ‘Intuitions’ and Scepticism about Judgment”, *Dialectica*, 58(1), 2004, pp. 109-155 (though Williamson is at pains to move away from talk of “intuition”).

want to consider one defense in particular, which we will call the “Reflection Defense”.

The Reflection Defense is made most explicitly by Antti Kauppinen. Here’s how Kauppinen puts the defense:

[W]hen philosophers claim that according to our intuitions, Gettier cases are not knowledge, they are not presenting a hypothesis about gut reactions to counterfactual scenarios but, more narrowly, staking a claim of how competent and careful users of the ordinary concept of knowledge would pre-theoretically classify the case in suitable conditions. The claim, then, is not about what I will call surface intuitions but about robust intuitions, which are bound to remain out of reach of the Survey Model of experimentalists, or so I will argue.⁷

Ernest Sosa also seems to have in mind something like the Reflection Defense when he writes, in response to the Experimentalists’ Challenge, that “[t]he upshot is that we have to be careful in how we use intuitions, not that intuition is useless”.⁸

According to this Reflection Defense, the hypothetical case intuitions studied by experimental philosophers are quite different from the kind of intuitions to which philosophers wish to appeal as evidence. Philosophers are interested in intuitions that are the product of careful reflection about the details of the hypothetical cases, about the key concepts involved in the cases, and so on. In contrast to these intuitions, the hypothetical case judgments studied by experimental philosophers consist in quick, surface-level judgments quite unlike the considered, thoughtful judgments that reflect attention to detail and careful consideration of hypothetical cases. As such, the hypothetical case intuitions studied by experimental philosophers aren’t relevant to the actual practice of appealing to hypothetical case intuitions as evidence and, therefore, their experiments leave that practice untouched and unscathed. The Reflection Defense thus aims to deflect the Experimentalists’ Challenge.

It is important to see that the Reflection Defense turns on an empirical hypothesis: intuitions of interest to philosophers—that is, thoughtful judgments based on reflection upon the case—won’t suffer from the same kinds of instability that seem to afflict the intuitions studied by experimental philosophers. Only if this empirical hypothesis is true will the Reflection Defense have any traction against the challenge. One way to test that hypothesis would be to look at some extant results in the experimental philosophy literature, and see what happens when one performs the same experiment on more reflective subjects. We chose to examine some recent findings suggesting that intuitions generated in response to

⁷ See, A. Kauppinen, “The Rise and Fall of Experimental Philosophy”, *op. cit.*, mss. pp. 6.

⁸ See, E. Sosa, “Experimental Philosophy and Philosophical Intuition”, *op. cit.*, pp. 105.

a version of Keith Lehrer's Truetemp Case are sensitive to whether, and which, other hypothetical cases are considered first.⁹ If the Reflection Defense is correct, then we should expect to find two things: first, people whose intuitions don't reflect thoughtful consideration of the cases will be subject to this effect; and, second, people whose intuitions do reflect attention to detail and careful consideration of hypothetical cases won't be subject to this effect.

The Instability of Truetemp Intuitions

A quick review of the previous study will be helpful. In his book, *Theory of Knowledge*, Keith Lehrer proposed a hypothetical case which he thought would be capable of generating an intuition that could be used as evidence of the falsity of epistemological reliabilism.¹⁰ Lehrer's hypothetical case, now commonly referred to as the "Truetemp Case", involves a person who has been caused to form true temperature beliefs by device which has been planted in his head. According to epistemological reliabilism, that the person's temperature beliefs are caused by a reliable cognitive process is enough to ensure that his temperature beliefs, when true, count as knowledge. According to Lehrer, however, when we consider the hypothetical case, we should be able to see that the person's temperature beliefs do not count as knowledge even though we accept that his beliefs are both true and caused by a reliable cognitive process. And, this intuition—that the person doesn't know—is supposed to count as evidence of the falsity of epistemological reliabilism: since epistemological reliabilism entails that the person does know and we have evidence (in the form of the intuition) that he doesn't know, we have evidence that epistemological reliabilism is false.

⁹ See, S. Swain, J. Alexander, and J. Weinberg, "The Instability of Philosophical Intuitions: Running Hot and Cold on Truetemp", *Philosophy and Phenomenological Research*, Vol. 76(1), (2008): pp. 138-155. These results were confirmed and replicated by Jen Wright who reported her results on the Experimental Philosophy Blog.

¹⁰ K. Lehrer, *Theory of Knowledge*, (Westview Press, 1990). Lehrer presents the case in the following way (pp. 163-164): Suppose a person, whom we shall name Mr. Truetemp, 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, call it a tempucomp, is implanted in Truetemp's 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. Now imagine, finally, that he 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, he thinks and accepts that the temperature is 104 degrees. It is. Does he know that it is?

But matters aren't this straightforward. Swain, Alexander, and Weinberg (op. cit.) found that just what intuition one will have in response to Lehrer's Truetemp Case actually depends on whether, and which, other hypothetical cases one has previously considered. Their subjects considered the following compressed version of Lehrer's case:

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

Subjects were then asked to indicate, using a five-point Likert scale (1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, and 5 = strongly agree), to what extent they agreed or disagreed with the following statement:

Charles knows that it is 71 degrees in his room.

They found that compared with subjects who considered the Truetemp Case before considering other cases (mean response = 2.8), subjects who were first presented with a clear case of knowledge were less willing (mean response = 2.4) to attribute knowledge in the Truetemp Case, and subjects who were first presented with a clear case of non-knowledge were more willing (mean response = 3.2) to attribute knowledge in the Truetemp Case ($p < 0.05$).¹¹

These results provided evidence that people's Truetemp intuitions are unstable. Proponents of the Reflection Defense, however, would reply that their results are evidence at best that *people who do not reflect on the case* have Truetemp intuitions that are unstable. In order to test the empirical commitments of the Reflection Defense, we gave subjects a version of the task used by the previous researchers, while in addition measuring our subjects' Need For Cognition (NFC) scores. This personality trait reflects the extent to which a person is intrinsically motivated to engage in sustained and effortful cognitive tasks. People with high NFC scores are more likely to have intuitions that reflect close attention to detail and careful consideration of the relevant content of hypothetical cases that they are presented with; people with lower NFC scores are more likely to have intuitions that reflect quick, surface judgments that don't reflect close attention to detail or careful consideration of the relevant content of hypothetical cases. Proponents of the Reflection Defense will thus make the prediction that higher NFC subjects will show a reduced or eliminated instability here; we, in the spirit of productive contrarianism, made the prediction that the higher NFC subjects would show the same instability. The surprising result: both predictions were wrong.

¹¹ An independent sample *t*-test was used.

Instability and Need For Cognition

Our subjects consisted of 160 students attending undergraduate classes at a large Midwestern university. Each subject was asked to participate in the study by completing a two-page survey. The first page included basic instructions and four hypothetical cases. Subjects were asked to indicate, for each hypothetical case, the extent to which they agreed or disagreed with a target statement attributing knowledge to the protagonist whose epistemic condition is described in the case using a 5-point Likert scale (1: strongly disagree, 2: disagree, 3: neutral, 4: agree, 5: strongly agree). The second page consisted of a standard survey used in social psychology to help determine a subject's NFC score together with several questions intended to help collect basic demographic data including age, gender, number of philosophy courses take, and standardized college admissions exam scores.

Subjects were randomly assigned to two different survey versions which differed only in the order in which the four hypothetical cases were presented. In addition to the Truetemp Case, each survey version contained three other hypothetical cases: a clear case of knowledge, a clear case of non-knowledge, and a version of Ginet's Fakebarn Case. In the clear case of knowledge, a chemist named Karen comes to believe that the mixing of two common floor disinfectants will create a poisonous gas on the basis of an article published in a leading scientific journal. In the clear case of non-knowledge, a person named Dave comes to believe that a coin will land heads on the basis of a "special feeling", where his decisions made on the basis of this "special feeling" are no better than chance.

The mean NFC score for our subjects was 54 with a standard deviation of 12. On this basis, we designated subjects with NFC scores ranging from 0-42 as "low-NFC", subjects with NFC scores ranging from 43-66 as "mid-NFC", and subjects with NFC scores ranging from 67-90 as "high-NFC". According to the hypothesis under consideration, we should expect the intuitions of subjects with low- and mid-NFC scores to be sensitive to the context in which the Truetemp Case is considered and the intuitions of subjects with high-NFC scores to be insensitive to the context in which the Truetemp Case is considered. What we found, instead, is that the intuitions of all subjects—regardless of their NFC scores—show sensitivity to the context in which the Truetemp Case is considered. More specifically, we found that, for low- and mid-NFC subjects, a person's willingness to attribute knowledge to Charles in the Truetemp Cases varied depending on the context in which the case was considered. Subjects who considered the case after a clear case of knowledge were less likely (mean response = 2.37) than those who considered the case after a clear case of non-knowledge (mean response = 2.89)

to attribute knowledge to Charles.¹² This result replicates what was found in the initial Swain, Alexander, and Weinberg study.

For the debate over the Reflection Defense, the interest is not in the low/mid-NFC subjects, but in the high-NFC subjects. And here our results caught us off guard. For high-NFC subjects, a person's willingness to attribute knowledge to Charles again varied depending on the context in which the case was considered – but in the opposite direction of the variation with the low/mid-NFC subjects! Subjects who considered the case after a clear case of knowledge were *more* likely (mean response = 2.64) than those who considered the case after a clear case of non-knowledge (mean response = 1.63) to attribute knowledge to Charles.¹³ We note that most of this flip in the direction of instability is driven by the high-NFC subjects' responses to the case where they saw a clear case of non-knowledge first.

The proponent of the Reflection Defense might try to object that we have not, after all, shown that our high-NFC subjects actually considered the cases *really, really* reflectively, and so it *might* still be the case that, if only they were to cross some sufficiently high threshold of reflectivity, then the instability would fade. But this line of response misconstrues the dialectical situation. The argument here is an empirical one, and thus trades on probabilities, and not mere possibilities. That a bit more reflection could make the SAW order effects go away was a claim that, in advance of our findings, had some rough empirical plausibility. That a *lot* of reflection can dissolve the effects is a claim that, in the face of our findings, is certainly conceivable, but not much more than that. The Reflection Defense will need its own empirically-ascertained contrary findings in order to be an argument in good standing.

The sheer *surprisingness* of this result underscores an important element of the Experimentalist Challenge: we just do not currently possess a very good sense of just where and how our intuitions about various sorts of cases may be susceptible to various sorts of instabilities. This is a key disanalogy between intuitions on the one hand and sense-perception on the other. As part of both our biological and social inheritance, we generally have a decent sense of where sense-perception may be led astray. It is far from perfect, as recent work on, e.g., eyewitness testimony has shown. But it is pretty good nonetheless, and is more than adequate for most normal applications of sense-perception. We can thus stress one way that the Experimentalists' Challenge is specific to intuitions, and does not necessarily need to plague other putative sources of evidence: it applies only to places where can expect instability to be a threat, but where we do not have the background knowledge or practices to quarantine that threat. The threat is

¹² An independent *t*-test revealed that the difference between these means is statistically significant at the level $p = 0.029$

¹³ An independent *t*-test revealed that the difference is statistically significant at the level $p = 0.037$.

mostly quarantined for sense-perception, and not at all for the kind of hypothetical case intuitions that the experimentalists are challenging.

A Possible Explanation

It is worth noting that social psychologists working on priming effects and need for cognition have reported results similar to ours, cases where high- and low-NFC subjects show inverted reactions to priming. For example, it has been reported that when primed with the concept of persistence, low-NFC subjects tend to form a more favorable impression of an ambiguously described individual than when primed with the concept of stubbornness while high-NFC subjects displayed the opposite pattern.¹⁴ Similar results have been reported with regard to emotional states and event estimations.¹⁵ When first asked to recall a time in which they were very sad low-NFC subjects tended to estimate more sad events in the future (e.g., more people will have to euthanize a pet this year) than when asked to recall a very happy or angry time. Again high-NFC subjects went the other way. So finding oppositely directed variation among low- and high-NFC subjects is not without precedent. These are but two examples.

Of course these two report results that are, in a way, different from ours. They found a greater tendency toward *assimilation* among low-NFC subjects and toward *contrast* among high-NFC subjects. We found the reverse direction of effects. For example we found that when first presented with a clear case of non-knowledge, relatively low NFC subjects tended to attribute knowledge to Truetemp Charles, rather than deny knowledge. But even the reverse can be explained, given the conditions of our survey. The most natural way to explain the above social-psychological results is to see the low-NFC subjects as providing the cognitively least demanding, or *default*, response and high-NFC subjects as giving the more demanding, *correction-attempted*, response. It's just that high-NFC subjects overcorrected.¹⁶ If that is right, then we could expect circumstances of flipped variation, just like what we found. It would require the default effect of the prime to be contrast (rather than assimilation).

¹⁴ See pp. 32-33 of L.L. Martin, J.L. Seta, and R.A. Crelia, "Assimilation and Contrast as a Function of People's Willingness and Ability to Expand Effort in Forming an Impression", *Journal of Personality and Social Psychology* 59(1), 1990, pp. 27-37.

¹⁵ See pp. 408-411 of D. DeSteno, R.E. Petty, D.T. Wegener, and D.D. Rucker, "Beyond Valence in the Perception of Likelihood: The Role of Emotion Specificity," *Journal of Personality and Social Psychology*, 78(3), 2000, pp. 397-416.

¹⁶ This is the explanation that the researchers themselves gave. See, for example, DeSteno, *et al.*, *op. cit.*, pp. 408-409. See also R.E. Petty, K.G. DeMarree, P. Briñol, J. Horcajo, and A.J. Strathman, "Need for Cognition Can Magnify or Attenuate Priming Effects in Social Judgment", *Personality and Social Psychology Bulletin*, in press.

There is experimental evidence demonstrating that the default could go that way. When subjects are briefly presented the name of an extremely athletic individual (in this case, ‘Michael Jordan’) they tend to give lower assessments of their own athletic abilities compared to the presentation of an extremely non-athletic individual (‘Pope John Paul’).¹⁷ Exactly what factors determine the default process is complicated, but evidence suggests that one tendency is this: when an extreme example precedes a task that requires evaluation of a particular, the default effect is contrast.¹⁸

Notice that the experimental the experimental design used here involved presenting subjects with either a clear case of knowledge or a clear case of non-knowledge before they saw the Truetemp Case. The social-psychological work on prime extremity suggests that these conditions make contrast the default response, especially when it comes to the very poor epistemic standing of Dave, the coin flipper. If so, borrowing only the basics from the social psychologists’ explanations, we can explain why relatively low NFC subjects will be more inclined to give the default, contrasting response, as found. This explanation works as long as the primes are blatant enough to push the lower NFC subjects one way and to trigger the high-NFC subjects’ correction processes. Clearly whether that is the case with our results, and the above results, is not yet settled. Only further empirical research will tell. But for now, our results are consistent with a number of extant psychological results, and such results may form a fruitful foundation to be built upon in further exploring the influence of need for cognition on people’s reported judgments.

Conclusion

The NFC study suggests that even intuitions that are likely to have involved attention to detail and careful consideration of hypothetical cases are unreliable. One upshot of this study is that it reduces the plausibility of the Reflection Defense. Since the kind of intuitions supposedly of interest to philosophers are as unstable – just *differently* unstable – as those studied by experimental philosophers, one can’t escape the challenge by appealing to such a difference. An additional upshot of this study is that it presents further problems for someone wishing to defend standard philosophical practice by appeal to epistemological contextualism.¹⁹ It seems like the contextualist might have a nice story to tell about the results of the Swain, Alexander, and Weinberg study. According to this story, what it is for a belief to count as knowledge depends, at

¹⁷ See pp. 691-993 of T. Mussweiler, K. Rüter, and K. Epstude, “The Man who wasn’t There: Subliminal Social Comparison Standards Influence Self-Evaluation,” *Journal of Experimental Social Psychology* 40, 2004, pp. 689-696.

¹⁸ For more on default contrast effects, see Mussweiler, *et al.*, *op cit.* and Petty, *et al.*, *op. cit.*

¹⁹ For further discussion of this kind of response to experimental philosophy, see S. Swain, J. Alexander and J. Weinberg, “The Instability of Philosophical Intuitions: Running Hot and Cold on Truetemp”, *op. cit.*

least in part, on the context the belief is being evaluated. In particular, in some contexts the standards a belief must pass in order to count as knowledge are higher than they are in other contexts. For the contextualist, then, it would not be at all surprising for the contextualist that when the Truetemp Case follows a clear case of knowledge, our intuition will be that Truetemp doesn't know. In such a context, the standards for attributing knowledge have been raised and Truetemp fails to meet those standards. Similarly, for the contextualist, we should expect that when the Truetemp Case follows a clear case of non-knowledge, our intuition will be that Truetemp knows. In such a context, the standards for attributing knowledge have been lowered and Truetemp meets those standards. In this way, contextualists can tell a story quite consonant with the empirical results of the Swain, Alexander, and Weinberg study. What's not clear, however, is that they can tell quite so neat a story concerning the empirical results of the NFC study. First, according to contextualism, the standards a belief must pass in order to count as knowledge are raised following a clear case of knowledge and lowered following a clear case of non-knowledge. How then would the contextualist explain that, for high-NFC people, Truetemp counts as knowing in high-standard contexts but doesn't count as knowing in low-standard contexts? Additionally, the NFC study would seem to indicate that Truetemp intuitions are sensitive not only to context but also to features about the person who is considering the case. And, it is not clear that contextualism has a story consonant with *this* kind of sensitivity.