**Knowing from the Armchair That Our Intuitions Are Reliable**

Ram Neta (University of North Carolina—Chapel Hill)

In recent years, a growing body of experimental literature has called into question the reliability of our intuitions about hypothetical cases, and thereby called into question the use of intuitions in philosophy. In this paper, I critically assess one prominent example of this challenge -- namely, Swain, Alexander, and Weinberg's recent study of order effects on the Truetemp intuition. I argue that the very data that Swain, Alexander, and Weinberg find do not undermine the reliability of intuition, but instead support it. I also show how intuition can itself be marshaled in the service of figuring out just when we can and cannot expect to find order effects on our intuitions.

Many philosophers assume that our intuitions – our spontaneous, seemingly non-inferential and non-perceptual inclinations to judge – as to whether or not a particular property is instantiated in a hypothetical case can provide us with evidence that bears on some philosophical hypothesis about the nature of that property.  For instance, our intuition to the effect that the property of knowledge is not instantiated in Gettier cases provides us with evidence against justified, true belief accounts of knowledge; our intuition to the effect that the property of causation is instantiated in pre-emption cases provides us with evidence against some counterfactual accounts of causation; and our intuition to the effect that the property of consciousness is not instantiated in a zombie case provides us with evidence against physicalist accounts of consciousness.  In each of these cases, and many others, philosophers commonly and confidently assume that, in deciding whether a particular property is instantiated in a hypothetical case, it is generally reasonable to accept – independently of any corroboration they may receive from other sources – the verdicts rendered by our own intuitions concerning whether the property in question is instantiated in the hypothetical case in question.  Let's call this assumption the assumption of*Intuitional Reliability*. If Intuitional Reliability is true, then we can reasonably rely upon our uncorroborated intuitions concerning whether a particular property is instantiated in a particular case to tell us whether the property really is instantiated in that case, and so to help us confirm or disconfirm hypotheses (about, say, the nature of that property) that make predictions about whether the property is instantiated in cases of a certain sort. (Note that I use the term “reliability” to denote a normative property, viz., the property of being such as can reasonably be trusted. While this normative property is not identical to any statistical property, e.g., truth-conduciveness, it is nonetheless true that Intuitional Reliability implies that intuitions are, at least in certain normal circumstances, for the most part true.)

A number of experimental philosophers have recently criticized the assumption of Intuitional Reliability.  More specifically, these experimental philosophers have aimed to show that a person's intuitions about the instantiation of some property in a hypothetical case vary significantly with factors that are obviously extraneous to the question of whether the property in question really is instantiated in that case.  In other words, these intuitions are like the readings of a device that is generally used to measure length, but the readings of which are highly sensitive to some factor that is independent of length, that varies significantly across normal circumstances (e.g., barometric pressure), and the variations of which are not detectable by use of the device itself.  Such a device would, of course, be a poor guide to the lengths of things under normal circumstances, and so too, it is suggested, our uncorroborated intuitions are poor guides to the properties the instantiation of which they judge.

Let's examine one recent and very prominent example of this form of challenge to Intuitional Reliability.  Stacey Swain, Joshua Alexander, and Jonathan Weinberg (henceforth, SAW) target the reliability of our intuitions in response to Keith Lehrer's "Truetemp" thought experiment, which is intended to serve as a counterexample to reliabilism.  Recall that Truetemp has, unbeknownst to him, had a tempucomp implanted in his brain.  (A tempucomp is a device that accurately reads the temperature and causes a true belief about that temperature to form spontaneously in the brain in which it is implanted.)  As a result, Truetemp has many true beliefs about the temperature, but he does not know what their source is or whether it is reliable.  Lehrer claims that, although Truetemp’s belief-forming process is reliable, the beliefs that he forms by means of this process are not knowledgeably held, and this is because Truetemp has no clue as to the source of those beliefs, or the reliability of that source.  Lehrer concludes that a reliable cognitive process cannot yield knowledge unless the believer is aware that the process is reliable. In other words, the mere fact that the process is reliable does not suffice, Lehrer concludes, to make the true beliefs that are formed by that process knowledgeably held.

This thought experiment is intended to serve as a counterexample to reliabilist theories of knowledge, and so to refute such theories.  But does it do so?  SAW have their doubts.  They begin their recent influential paper "The Instability of Philosophical Intuitions:  Running Hot and Cold on Truetemp," with the following paragraph:

A growing body of empirical literature challenges philosophers' reliance on intuitions as evidence based on the fact that intuitions vary according to philosophically irrelevant factors, such as cultural and educational background or affective state.  Our research extends this challenge, demonstrating that intuitions vary according to whether, and which, other thought-experiments are considered first.  We critique the use of intuitions, such as those generated by Lehrer's Truetemp Case, as evidence, on the grounds that intuitions about this case are easily manipulated:  compared to subjects who receive the Truetemp Case first, subjects first presented with a clear case of knowledge are less willing to attribute knowledge in the Truetemp Case, and subjects first presented with a clear case of non-knowledge are more willing to attribute knowledge in the Truetemp Case.[[1]](#footnote-1)

SAW's data are not altogether surprising.  But their discussion of the data is fascinating, and deserves extended response.  The present paper is intended as a response.

In section I, I describe their data, state the conclusion that they draw from these data, and consider an objection that Ernest Sosa makes against their conclusion, and their reply to Sosa’s objection.  SAW's reply to Sosa depends upon their claim that there is an important difference between their appeal to order effects to call into question our reliance on intuitions in philosophy, on the one hand, and a seemingly analogous but obviously unsuccessful appeal to order effects to call into question our reliance on perception in ordinary life, on the other.  But is this claim true?  To answer this question, I begin by considering (in section II) some of the epistemological intuitions that SAW themselves seem to rely on in their own discussion of their data.  It will turn out that, on the most charitable interpretation that I can find, SAW's own discussion of their data presupposes a substantive (and intuitively plausible) thesis concerning the property of knowledge.  This is the thesis that someone can fail to have knowledge that *p* because of substantial evidence that she does not possess against the reliability of the cognitive ability that she exercises in believing that *p*:  let's call this the "no substantial defeaters" thesis.  As I argue in section III, this "no substantial defeaters" thesis has an important implication: given that Lehrer's Truetemp story involves a defeater (there is some misleading evidence, not possessed by Truetemp, against the reliability of Truetemp’s cognitive ability), and given that this defeater is neither clearly substantial nor clearly insubstantial, and given that whether we think of it as substantial or not will naturally depend on what other defeaters we are contrasting it with; given all this, it is intuitively plausible that our intuitions about Lehrer's Truetemp case will display just the sorts of order effects that SAW find in their data.  In other words, I will argue, our intuitions about order effects predict that our intuitions about Truetemp will display the very order effects that SAW find. And so if Intuitional Reliability is true, it follows that our intuitions about Truetemp very likely do display these order effects.  (Of course, our intuitions do not predict anything quite as precise as SAW's quantitative data.  But then nothing in SAW's reasoning turns on the precise quantitative features of their data either.)  In sum, SAW's own discussion of their data tacitly presupposes a substantive and intuitively plausible epistemological assumption, but that intuitively plausible assumption makes SAW's data intuitively predictable.  And so SAW's data cannot be used to criticize the assumption of Intuitional Reliability, since Intuitional Reliability predicts the very same data that SAW find.  Or, to state the point with a finer grain, the conditional probability of SAW's data on the hypothesis that Intuitional Reliability is true is at least as high as the conditional probability of SAW's data on the hypothesis that Intuitional Reliability is false, and so SAW's data cannot disconfirm Intuitional Reliability.

Here is a way to sum up the argument of this paper. SAW’s argument runs as follows:

(SAW 1) Intuitions about cases suffer from order effects (e.g., Truetemp).

(SAW 2) If intuitions about cases suffer from order effects, then Intuitional Reliability is not true.

(SAW 3) Therefore, Intuitional Reliability is not true.

This paper does not attempt to show that (SAW 3) is false. Rather, it attempts to call into question the soundness of that argument by calling into question SAW’s reason for believing (SAW 2). And here is how the argument of this paper goes:

(1) It is intuitive that *S* knows that *p* only if *S* has no substantial defeaters of the reliability of the cognitive competence that *S* exercises in believing that *p*.

(2) It is intuitive that the notion of *substantial defeat* is vague, and that a defeater can be more or less substantial.

(3) It is intuitive that the Truetemp case involves a borderline case of substantial defeat (Truetemp has a defeater of the reliability of his competence, but it's not a very substantial defeater).

(4) It is intuitive that, for a psychologically salient property X (e.g., how substantial a defeater is), when a borderline case of X is compared against clear cases of X, it seems less X-ish, and when a borderline case of X is compared against clear cases of non-X, it seems more X-ish.

(5) What is intuitive is, in normal circumstances, by and large true. (Intuitional Reliability)

(6) So, 1, 2, 3, and 4 are all probably true. (From 5)

(7) Therefore (probably), when the Truetemp case is compared against a clear case of substantial defeat (like Coinflip Dave), it seems less like a case of substantial defeat (and so more like a case of knowledge), and it is probably true that when the Truetemp case is compared against a clear case of no substantial defeat (like Karen the Chemist) it seems more like a case of substantial defeat (and so less like a case of knowledge).

Since (7) undermines SAW’s reason for believing (SAW 2), it follows that if the argument of this paper is sound, then SAW’s argument for (SAW 3) is unsuccessful.

Before proceeding, I must anticipate one very common line of resistance to my argument, as I’ve rendered it above. Premises 1 – 4 all begin with the locution “it is intuitive”. But, you might wonder, how precisely should this be understood? Should I be understood as claiming that something is intuitive to me? If so, then what is the interest of such a claim? Why should anyone care about what is intuitive to me? Should I be understood instead as claiming that something is intuitive to everyone? If so, then premises 1 – 4 are all demonstrably false.

Neither of the suggested options captures my intended meaning. To understand what I mean by the locution “it is intuitive”, think for a moment about the debate between realists and anti-realists in the philosophy of science concerning the relation between theory and observable data. Such debates all employ a concept of the observable, and take some things to fall clearly into the extension of the concept (e.g., tables and chairs) and other things to fall clearly into its anti-extension (e.g., electrons). But suppose we were to ask the parties to this dispute “what do you mean by ‘observable’? Do you mean observable to you personally? Or do you mean observable to all people?” Neither of these answers will do: nothing is observable to all people, since some people are blind and deaf; and no distinction of any philosophical importance turns on what is observable to you personally. Still, there seems to be something important at stake in the debate between realists and anti-realists in the philosophy of science, and so some substantial distinction to be drawn between the observable and the unobservable, even if the notion of the observable cannot be spelled out in either of the two ways envisioned. A more promising way to spell out the notion of the observable would be by appeal to normal human powers of observation, though it’s not at all clear just how to spell out the details of this proposal. Similarly, I claim, we should spell out the notion of the intuitive (employed in premises 1 – 4 of my argument) by appeal to normal human powers of intuition, and let the details of this proposal to be spelled out by analogy with the details of the proposal above concerning the observable – however precisely that goes.

Finally, I should note that SAW’s data are just the tip of an iceberg of data that have been collected on Truetemp and other cases since the publication of SAW’s paper. Since my focus in this paper is on a particular methodological issue that SAW raise, I will not be discussing the profusion of further data that have been gathered since the publication of their paper.[[2]](#footnote-2)

**Section I:  SAW's data and discussion**

So first, what are SAW's data?  SAW surveyed 220 students attending undergraduate classes at Indiana University, Bloomington, and assigned each of them randomly to one of eight conditions.  The conditions differed only in respect of the order in which the following four vignettes and questions were presented.  (Each of the eight conditions involved asking students to answer questions about all four of the vignettes.)

"We are investigating what different people's opinions are about knowledge.  In each question, please indicate to what extent you agree or disagree with that statement.

1.  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.  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.

Please indicate to what extent you agree or disagree with the following claim:  'Dave knew that the coin was going to land heads.'
\_\_\_Strongly agree   \_\_\_Agree   \_\_\_Neutral   \_\_\_Disagree   \_\_\_Strongly disagree

2.  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.

Please indicate to what extent you agree or disagree with the following claim:  'Charles knows that it is 71 degrees in his room.'
\_\_\_Strongly agree   \_\_\_Agree   \_\_\_Neutral   \_\_\_Disagree   \_\_\_Strongly disagree

3.  Suzy looks out the window of her car and sees a barn near the road, and so she comes to believe that there's a barn near the road.  However, Suzy doesn't realize that the countryside she is driving through is currently being used as the set of a film, and that the set designers have constructed many fake barn facades in this area that look as though they are real barns.  In fact, Suzy is looking at the only real barn in the area.

Please indicate to what extent you agree or disagree with the following claim:  'Suzy knows there is a barn near the road.'
\_\_\_Strongly agree   \_\_\_Agree   \_\_\_Neutral   \_\_\_Disagree   \_\_\_Strongly disagree

4.  Karen is a distinguished professor of chemistry.  This morning, she read an article in a leading scientific journal that mixing two common floor disinfectants, Clean 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!'

Please indicate to what extent you agree or disagree with the following claim:  'Karen knows that mixing these two products creates a poisonous gas.'
\_\_\_Strongly agree   \_\_\_Agree   \_\_\_Neutral   \_\_\_Disagree   \_\_\_Strongly disagree"

We can label each of the vignettes with the first letter of its protagonist's name, thus:  D, C, S, and K.  And so we can label each of the eight conditions into which subjects were randomly assigned by arranging these four letters in different orders, corresponding to the order in which the vignettes were presented in that condition.  Using this labeling scheme, we can represent SAW's data as follows:

Condition                Mean response to case C (Strongly agree = 5; Strongly disagree = 1)
KCDS                     2.4
CSDK                     2.6
SCDK                     2.9
DSCK                     2.9
CDSK                     3.0
CKDS                     3.0
DCSK                     3.2
DSCK                     3.6

In each condition, two standard deviations around the mean extended approximately .5 in each direction.  Clearly, subjects' responses to the question whether Charles knows that it is 71 degrees in his room depend, to some significant extent, on the order in which the vignettes are presented.  Subjects' responses concerning Charles depend significantly on whether case C is immediately preceded by case K (the case of chemist Karen) or by case D (the case of coinflip Dave).  Since the story about Charles is, in all philosophically important respects, just like Lehrer's story about Truetemp, it follows that SAW empirically demonstrate that order effects significantly affect the Truetemp intuition.  (As should be clear from the presentation of the data above, SAW's findings are actually much more quantitatively precise than anything captured by the non-quantitative summary statement I just gave.  But, for the purposes of the discussion to follow, this quantitative precision will not matter.)

Do these findings disconfirm Intuitional Reliability, and thereby undermine the use of intuitions as evidence in philosophy?  SAW critically examine a particular argument, offered by Ernest Sosa, to the effect that they do not do so.[[3]](#footnote-3)  Sosa points out that, just as our intuitions about cases are unreliable under certain conditions, so too are our perceptual judgments unreliable under certain conditions.  Sosa then writes:

One would think that the ways of preserving the epistemic importance of perception in the face of such effects on perceptual judgments would be analogously available for the preservation of the epistemic importance of intuition in the face of such effects on intuitive judgments.  The upshot is that we have to be careful in how we use intuition, not that intuition is useless. (p. 105)

To evaluate Sosa's suggestion, we must consider what are "the ways of preserving the epistemic importance of perception in the face of such effects on perceptual judgments", and then see whether or not these ways are indeed "analogously available for the preservation of the epistemic importance of intuition in the face of such effects on intuitive judgments."  So first, what, in general, are "the ways of preserving the epistemic importance of perception in the face of such effects on perceptual judgments"?  Roughly, they involve appeal to the distinction between conditions under which perception operates reliably and conditions under which it does not.  The boundary between the former set of conditions and the latter is not simple, but its contours are reasonably well known.  When the illumination is too dim or too bright, vision is neither precise nor accurate.  When a sound is too quiet, it is hard to discriminate its timbre.  And so on.  We know, by and large, under what conditions perceptual judgments are reliable and under what conditions they are not.  The fact that they are unreliable under some conditions has no tendency to show that such judgments are in general unreliable.

So, are these considerations "analogously available for the preservation of the epistemic importance of intuition" in the face of SAW's data?  SAW claim that they are not, and this is for a simple reason:  we are insufficiently familiar with the location of the boundary between those conditions under which intuition operates reliably and those conditions under which it does not.  As SAW put the point:

At this time, we don't know what is the parallel for intuition of making sure that the light is on; that is, we do not know which are the circumstances that render intuition reliable or unreliable.  With perception, by contrast, we are aware of the dimensions of variance and how to compensate for them.  We know how to turn out heads toward the speaker if we cannot hear well, or to squint if we are trying to read a distant road sign, or to cleanse the palate before evaluating a fine wine.  What our research indicates is that we do not have analogous knowledge relating to our practice of relying on intuitions. (p. 148)

What I'd like to do now is call into question SAW's claim in this paragraph.  Specifically, I will argue that, while we might be *much more* familiar with the range of circumstances in which perception is reliable than with the range in which intuition is reliable, we are sufficiently familiar with the latter that we can reasonably judge that such circumstances do not obtain in SAW's Truetemp experiment.  My argument for this last claim will proceed in several steps, but it will begin from consideration of some things that SAW themselves go on to say in their paper, immediately following their discussion of Sosa's argument.  That is the part of SAW's discussion to which I turn in section II.

**SECTION II:  Why SAW think that Sosa cannot reasonably dismiss their data**

In summarizing their reply to Sosa, SAW write the following:

As the empirical case against particular intuitions expands, it falls on those philosophers who wish to continue to employ intuitions as evidence to demonstrate that intuitions about their favorite thought-experiments are not susceptible to this, and other, problematic effects.  Sosa may be right that the upshot of the instability of intuitions is that we ought to be careful; the problem is that it is not clear what it means to be careful when it comes to intuition.  Thus, it falls on those philosophers who wish to continue relying on intuitions to figure out what it means to be careful.  We do not rule out *a priori* the possibility that they will be able to do so.  What we rule out is the possibility that they will be able to do so *a priori*. (pp. 148-9)

In the remainder of this paper, I will raise an objection to SAW's claim, made at the very end of this quoted passage, that we cannot figure out a priori what it means to be careful when it comes to intuitions.  But, in order to articulate that objection, let me first raise a question about something that SAW say earlier in the paragraph.  They say "as the empirical case against particular intuitions expands, it falls on those philosophers who wish to continue to employ intuitions as evidence to demonstrate that intuitions about their favorite thought-experiments are not susceptible to this, and other, problematic effects."  But what is the normative force of the point that this burden "falls on" those philosophers?  Suppose that those philosophers fail to discharge this supposed burden:  what, then, is SAW's charge against them?  Precisely what, according to SAW, are these philosophers guilty of?  Let's consider various interpretive options:

Is it that these philosophers are guilty of *falsely* assuming that intuitions about their favorite thought-experiments are not susceptible to order effects?  Certainly this cannot be what SAW want to say.  My failure to show that intuitions about my favorite thought-experiment are not susceptible to order effects of course does not imply that those intuitions actually are susceptible to order effects.

Is it that these philosophers are guilty of assuming *without any evidence* that intuitions about their favorite thought-experiments are not susceptible to order effects?  This cannot be what SAW want to say either.  It is extremely implausible that philosophers would make such an assumption without any evidence at least *from their own impressions of their own case*.  If a philosopher believes that intuitions about a particular thought-experiment are not susceptible to order effects, it is plausible that, at the very least, she has not noticed any such order effects in her own responses to the thought-experiment.  But the conditional probability of her not noticing any such order effects in her own case, on the hypothesis that there are no such order effects, is higher than the conditional probability of her noticing any such order effects in her own case, on the hypothesis that there are such order effects.  In other words, one's not noticing any such order effects is at least some evidence that there are no such order effects:  if there were such effects, it's less likely that one would not have noticed them.

So perhaps what SAW want to say is that these philosophers are guilty of assuming *contrary to their total evidence* that intuitions about their favorite thought-experiment are not susceptible to order effects?  But again, this cannot be what SAW want to say:  one way to insure that the assumption in question is consistent with our total evidence is to avoid coming into contact with any evidence that runs contrary to the assumption.  A more common way to insure the same result would be to reason from the obvious truth of certain judgments (e.g., that Gettier cases are not cases of knowledge) to the conclusion that, however the experimental data were collected, those data must be misleading. (This latter strategy was quite popular in response to the Weinberg, Stich, and Nichols data concerning Gettier cases, and it’s not obvious what, in general, is wrong with the strategy.[[4]](#footnote-4) If I hear that a recent paper published in the Journal of Philosophy defends the claim that no belief is ever more justified than any other, I can be rationally very certain that there is something wrong with this argument, even before I find out exactly how the argument goes.)

Perhaps, then, what SAW want to say is that these philosophers are guilty of assuming *without sufficient evidence* that intuitions about their favorite thought-experiments are not susceptible to order effects?  This proposal is not implausible but rather incomplete, for we may ask:  if SAW want to say that their philosophical opponents are guilty of assuming, *without sufficient evidence*, that intuitions about their favorite thought-experiments are not susceptible to order effects, then precisely what is it that such evidence is (according to SAW) not sufficient for?  Is it that such evidence is (according to SAW) not sufficient for reasonably believing that such intuitions are not susceptible to order effects?  But again, unless they want to set standards for reasonable belief at an unreasonably high level, SAW cannot plausibly claim this.  If a philosopher has not noticed any order effects in her own case, and she finds it so obvious that there are no order effects that she can reasonably dismiss evidence that purports to tell against her as misleading, then why shouldn't she believe that there aren't any such order effects?  Consider, for instance, Russell's famous "stopped clock" case, in which one sees a clock that one falsely believes to be working properly.  The clock, let us suppose, is stuck reading Noon, and one is looking at the clock at Noon and so forms the true belief that it is Noon.  Here, it is my strong intuition that one has a true belief that it is Noon, but one does not know that it is Noon (at least not if one has no source of information about the time that is independent of the clock).  I do not notice that this intuition of mine is subject to any order effects, I know of no evidence to suggest that the intuition is subject to order effects, and if someone were to tell me that there is some such evidence, I could reasonably dismiss this supposed evidence as misleading.  It seems clear that, without gathering any additional evidence, I can already reasonably (albeit defeasibly) believe that this intuition is not subject to order effects.  It also seems clear that I can, without gathering any additional evidence, reasonably (albeit defeasibly) proceed to theorize on the presupposition that the intuition is not subject to order effects.  If SAW want to deny these claims, then their denial is going to have to be based on some revisionary view about which beliefs are reasonable.  So, *if* my intuition about Truetemp intuitively seems to be as stable as my intuition about the stopped clock case (a big "if", and one that I will consider in some detail below), then SAW cannot plausibly claim that philosophers cannot reasonably believe, or proceed on the assumption that, their own Truetemp intuitions are not subject to order effects.

So how should we understand the claim that SAW are making in the passage I've quoted above?  We should not understand them as saying that, when philosophers assume that intuitions about their favorite thought-experiments are not susceptible to order effects, they are doing so falsely, or without any evidence, or contrary to their total evidence, or without sufficient evidence to make their assumption reasonable to hold or to act on.  So what should we understand SAW to be saying then?  The only plausible interpretive option that I can see is the following.  SAW are, I suggest, best understood as claiming the following:

When philosophers assume that intuitions about their favorite thought-experiments are not susceptible to order effects, they are assuming something that they do not know to be true.

This is an accusation which, it seems to me, SAW can quite plausibly level.  But why is this accusation, unlike the other accusations considered above as possible interpretations of SAW, plausible?  It's because knowledge, unlike other epistemic statuses (e.g., reasonableness), can be defeated by evidence of which one is entirely unaware. In particular, knowledge, unlike other epistemic statuses, can be defeated by evidence against the reliability of one’s own belief-forming faculties, even if one is entirely unaware of that evidence.  One can sustain the reasonableness of one's assumptions by failing to be notified of evidence against the reliability of one’s cognitive faculty, but one cannot sustain one's knowledge by clinging to ignorance.  This, recall, is what we called the "no substantial defeaters" hypothesis at the beginning of this paper.

I conclude, then, that if we are to find something plausible for SAW to mean in the quoted passage above, it is just this:

When philosophers assume that intuitions about their favorite thought-experiments are not susceptible to order effects, they are assuming something that they do not know to be true.

How do SAW's data support this claim?  How does the empirical evidence of order effects in our intuitions support the idea that, when philosophers assume that intuitions about their favorite thought-experiments are not susceptible to order effects, they are assuming something that they do not know to be true?  This is because SAW's own data (they claim) provide evidence against the assumption that intuitions about philosophical thought-experiments are not susceptible to order effects.  But such evidence is precisely the sort of thing that, according to the no substantial defeaters hypothesis, defeats knowledge that intuitions about philosophical thought-experiments are not susceptible to order effects.  So philosophers don't know their assumption to be true because SAW's data -- and other data like it -- provide evidence against the truth of their assumption, and such evidence robs them of knowledge whether or not they are aware of that evidence, and whether or not their assumption is true.

Here, we have located a plausible claim for SAW to making about the significance of their data, and we have found that this claim presupposes the very plausible no substantial defeaters hypothesis about knowledge.  But notice:  if the no substantial defeaters hypothesis is true, then we should expect the following to be true of the four cases given in SAW's questionnaire.  First, the case of Coinflip Dave is a very clear case of non-knowledge.  Although Dave might correctly and confidently believe of a particular toss that it landed heads, this belief cannot be knowledgeably held, because there is very strong evidence (perhaps not possessed by Dave) against the reliability of Dave's belief-forming method, viz., that people in general do no better than guessing at forecasting the results of coin tosses, and that Dave has no reason to suppose that he does any better than other people do.  Second, the case of Karen the Chemist is a clear case of knowledge, since the case does not suggest any evidence whatsoever that might tell against the reliability of her way of forming a belief in the toxic effects of mixing Clean Plus and Washaway.  Third, the case of Truetemp Charles is neither a clear case of knowledge nor a clear case of non-knowledge.  While Charles correctly and confidently forms a belief about the precise ambient temperature, and while he is reliably right in such beliefs, there is nonetheless some evidence that tells against the reliability of his way of forming such beliefs, viz., that people in general are not reliably right about the precise ambient temperature, and that Charles has no very good reason (apart, perhaps, from the unusual precision in the content of his belief) to suppose that he does any better than other people do.  Notice that this is the same general kind of evidence that tells against Coinflip Dave's belief about the result of the coin toss:  in each case, there is evidence about how reliable human beings are in general when they form beliefs in a certain spontaneous way about a given subject matter.  Coinflip Dave has no reason whatsoever to believe that he does any better than other human beings do, and Truetemp Charles has little reason to believe this.  Finally, the case of Suzy is neither a clear case of knowledge nor a clear case of non-knowledge.  While Suzy correctly and confidently believes of a particular object that it is a barn, and while she also has reason to believe that her beliefs about barns are reliably formed in this environment (given that she doesn't realize that there's anything unusual about the countryside through which she is driving), there is nonetheless some evidence that tells against her belief, viz., that her beliefs about barns are not reliably formed in this environment.  Notice, by the way, that this is not the same kind of evidence that tells against Coinflip Dave's belief about the result of the coin toss, or Truetemp Charles's belief about the precise ambient temperature.  The evidence that tells against Dave's belief or Charles's belief is evidence about what human beings are generally reliable at detecting, whereas the evidence that tells against Suzy's belief is evidence about the odd features of her environment.

So, if we accept the very plausible no substantial defeaters hypothesis, then we should judge that the case of Coinflip Dave is a clear case of non-knowledge, the case of Karen the Chemist is a clear case of knowledge, and the cases of Suzy and Truetemp Charles are neither (and for different reasons).  And these verdicts about the cases all seem more or less plausible.  In the next section, we will see that, just as there are powerful intuitions that favor these verdicts about the four cases, so too there are powerful intuitions that make it plausible that our intuitions about the Truetemp case will display precisely the kinds of order effects that SAW find, and that our intuitions about the Suzy case will not display the same kinds of order effects.  This point will be important because, if it is true, then Intuitional Reliability strongly predicts the very same order effects that SAW find, and so the order effects cannot be used as evidence against Intuitional Reliability.

**SECTION III:  Intuitions about contrast and order effects**

Let me now describe three different hypothetical situations, and ask you, the reader, to make a judgment about whether or not a particular property is instantiated in each of the situations.  In order for me to describe these situations, I must first begin by mentioning three different vignettes, each of which concludes with a question:

Vignette 1.  Karen casually mentions to John that she was planning to grab lunch this afternoon at Ahmed's Deli.  "If you'd like to join me, I'll be there at half past Noon," she says to him.  "Well, uh, yeah... I… I’ll be there," he replies, in a very tentative, uncertain tone of voice.

Please indicate to what extent you agree or disagree with the following claim:  'John promised Karen that he would meet her at Ahmed's Deli at half past Noon.'
\_\_\_Strongly agree   \_\_\_Agree   \_\_\_Neutral   \_\_\_Disagree   \_\_\_Strongly disagree

Vignette 2.  Karen tells John that she'd like to talk to him about something this afternoon, and asks if he'll meet her to discuss it over lunch at Ahmed's Deli.  "Can we meet there at half past Noon?" she asks him.  "Yes, I'll be there," he replies with some conviction.

Please indicate to what extent you agree or disagree with the following claim:  'John promised Karen that he would meet her at Ahmed's Deli at half past Noon.'
\_\_\_Strongly agree   \_\_\_Agree   \_\_\_Neutral   \_\_\_Disagree   \_\_\_Strongly disagree

Vignette 3.  Karen tells John that she is being harassed by her boss, and she needs John to videotape the harassment so that she can present evidence to the management of her company.  "My boss is meeting me for lunch at Ahmed's Deli today at half past Noon," she says to him.  "I absolutely need you to be there to videotape our interaction.  Can you promise me that you'll be there?"  "Absolutely," John replies.  "I guarantee it.  I'll be there:  you can count on it."

Please indicate to what extent you agree or disagree with the following claim:  'John promised Karen that he would meet her at Ahmed's Deli at half past Noon.'
\_\_\_Strongly agree   \_\_\_Agree   \_\_\_Neutral   \_\_\_Disagree   \_\_\_Strongly disagree

Now, having described these three vignettes and questions, I can proceed to describe the two hypothetical situations that I want you to consider.

Hypothetical situation 1.  Jones is a normal English-speaking adult, who has agreed to participate in a brief experiment.  He is given vignettes 1, 2, and 3, in that order, and asked to answer each question after reading the vignette that precedes that question.

Now, I ask you:  roughly what answers does Jones give to the three questions? If you are like me, you will intuitively make the following judgments about Jones:  First, Jones disagrees that (or is, at best, neutral as to whether) John promised in vignette 1.  Second, Jones strongly agrees that John promised in vignette 3.  And finally, Jones's response to vignette 2 is somewhere between these points, though closer to agreement than to neutrality.  Those are all intuitively plausible judgments about Jones in hypothetical situation 1.

Now, consider:

Hypothetical situation 2.  Jones is a normal English-speaking adult, who has agreed to participate in a brief experiment.  He is given vignettes 3, 2, and 1, in that order, and asked to answer each question after reading the vignette that precedes that question.

Now, I ask you:  roughly what answers does Jones give to the three questions? If you are like me, you will intuitively make the following judgments about Jones:  First, Jones agrees (or perhaps strongly agrees) that John promised in vignette 3.  Second, Jones disagrees that (or is, at best, neutral as to whether) John promised in vignette 1.  And finally, Jones's response to vignette 2 is somewhere between these points, though closer to neutrality than to agreement.  Those are all intuitively plausible judgments about Jones in hypothetical situation 2.

If Jones starts off with vignette 1 and then moves to vignette 2 and vignette 3 in that order, it's plausible that Jones will treat vignette 2 as a reasonably clear case of John's promising, and treat vignette 3 as an utterly obvious case of John's promising.  If, however, Jones starts off with vignette 3 and then moves to vignette 2 and vignette 1 in that order, it's plausible that Jones will treat vignette 2 as a somewhat borderline case of John's promising, and treat vignette 1 as a reasonably clear case of John's not promising.  In short, it is intuitive that there will be an order effect in Jones's answer to the question that follows vignette 2.  We could, of course, empirically test whether or not such an order effect is actually found in some population.  But, whatever the results of this empirical test would be, my present point is simply this:  it is intuitive (whether or not it is true) that such an order effect will be found in a normal individual.  I am not saying that intuition is right to tell us this; I am saying only that this is what it tells us.

In general, it is intuitively plausible that the order in which cases are presented can, at least in some cases, affect a normal person's judgments about those cases.  One way in which it is prone to do so is by using contrast to highlight certain features of those cases.  There is a clear contrast between vignette 1 and vignette 2, and another clear contrast between vignette 2 and vignette 3, and the features of these cases that make them contrast with each other are highlighted when the cases are presented in order.

But what features are these, that are thus highlighted?  One simple hypothesis is this:  the one feature of these cases that is highlighted by dint of their contrast with each other is the degree to which each case is a clear case of promising.

This simple hypothesis, however, cannot be right. For consider the following vignette:

Vignette 4.  Karen tells John that she'd like to talk to him about something this afternoon, and asks if he'll meet her to discuss it over lunch at Ahmed's Deli.  "Can we meet there at half past Noon?" she asks him.  John, who is very drunk, replies "Abshoholutely!  I wishimoshadefini... I promish I will.  I promish I be there.  Mosh definite.  Promish.  Crosh my heart."

Please indicate to what extent you agree or disagree with the following claim:  'John promised Karen that he would meet her at Ahmed's Deli at half past Noon.'
\_\_\_Strongly agree   \_\_\_Agree   \_\_\_Neutral   \_\_\_Disagree   \_\_\_Strongly disagree

In vignette 4, just as in vignette 1, we have a case that is not at all a clear case of promising. But it is not plausible that putting vignette 4 before or after vignette 2 would result in any significant order effects. This implies that our simple hypothesis concerning the contrast between stories 1, 2, and 3 cannot be right. A better and more plausible explanation for the order effect that would plausibly be found in presenting cases 1, 2, and 3 is rather this: issuing a promise to F without explicitly saying that one is issuing that promise involves *assuring* one’s interlocutor that one will F. Thus, only in so far one assures one’s interlocutor that one will F can one promise to F without explicitly saying “I promise”. Since the degree of assurance that John offers Karen in vignette 1 is very low, the assurance that he offers Karen in vignette 2 seems strong by contrast when vignette 2 is presented after vignette 1. And since the degree of assurance that John offers Karen in vignette 3 is very high, the assurance that he offers Karen in vignette 2 seems weak by contrast when vignette 2 is presented after vignette 3. This, at any rate, is what our intuitions predict. And the hypothesis that explains the intuitively predicted order effect is itself intuitively plausible.

Why have I considered these four vignettes about promising, and the order effects that intuition predicts we would find among them? I hope it is clear that these four vignettes are very closely analogous to the four vignettes that SAW consider (those of Coinflip Dave, Truetemp Charles, Chemist Karen, and Façade Suzy). And just as there is an intuitively plausible explanation of the intuitively predictable order effects among our vignettes about promising, so too is there an intuitively plausible explanation of the intuitively predictable order effects among the cases that SAW consider. The latter explanation relies on a principle that, as I have argued above, SAW themselves assume, viz., the no substantial defeaters hypothesis. The explanation goes like this: As the no substantial defeaters hypothesis claims, knowledge can be defeated by evidence against the reliability of one’s belief-forming faculty, even if one is unaware of that evidence. Now, in the case of Coinflip Dave, there is compelling evidence against the reliability of Dave’s way of forming his belief that the coin will land heads, namely: such special feelings as the one that Dave has are, in general, completely unreliable as indicators of whether a coin will land on a particular side. In the case of Truetemp Charles, there is some evidence against the reliability of Charles’s way of forming his belief that the temperature is 71 degrees, namely: although peoples’ spontaneous but rough estimations of temperature are normally reliable, their spontaneous but precise estimates tend to be much less reliable. In the case of Chemist Karen, there is no evidence at all against the reliability of her way of forming her belief that mixing the two cleaning products will create a poisonous gas. So the cases of Coinflip Dave, Truetemp Charles, and Chemist Karen vary along a single dimension concerning how substantial are the defeaters against the reliability of the subject’s belief-forming faculty, with the case of Coinflip Dave being at one end of the spectrum (the defeater is very substantial), the case of Chemist Karen being at the other end (there is no defeater at all), and the case of Truetemp Charles being intermediate between them. This would explain why it is that contrasting the case of Truetemp Charles with the case of Coinflip Dave would make the case of Truetemp Charles seem more like a case of knowledge, while contrasting the case of Truetemp Charles with the case of Chemist Karen would make the case of Truetemp Charles seem less like a case of knowledge. And this is precisely what SAW found.

This would also explain why no order effect is found with the case of Façade Suzy: while that case is neither a very clear case of knowledge nor a very clear case of non-knowledge, what makes it a borderline case is not a matter of there being evidence against the reliability of Suzy’s cognitive faculty, but rather a matter of the peculiarities of Suzy’s environment. So the case of Façade Suzy – unlike the case of Truetemp Charles – does not occupy an intermediate position on the spectrum that runs from the case of Chemist Karen to the case of Coinflip Dave. If we want to see order effects on our intuitions about the case of Façade Suzy, we will have to contrast it with a case in which there are no facades in Suzy’s environment, and then contrast it with a case in which there are almost exclusively facades in Suzy’s environment, and Suzy is looking at the only real barn anywhere. It is intuitively plausible that contrasting the case of Façade Suzy with those other cases would result in order effects. But intuitions are silent on the issue of how significant those order effects would be.

I anticipate an objection: If it is so intuitive that our intuitions about Truetemp will be subject to order effects, then why was did analytic epistemologists like Lehrer – epistemologists who operate by attending to their intuitions – take Truetemp to constitute a counterexample to reliabilism? Why did they not instead treat Truetemp as an unclear case of knowledge? My answer to this is simple: these epistemologists did not treat Truetemp as an unclear case of knowledge precisely because they were thinking of it in contrast to clear cases of knowledge like Karen the Chemist. And when compared to such cases, Truetemp really does not look like a case of knowledge. (That, indeed, is part of what I am claiming in this paper.) So what Lehrer and other analytic epistemologists noticed was that, compared to cases like that of Karen the Chemist, the Truetemp case really does not look like a case of knowledge. But there is a further fact that Lehrer and others did not notice, which is that, compared to cases like that of Coinflip Dave, the Truetemp case really does look like a case of knowledge. So it’s not that Lehrer and others were wrong about what was intuitive; rather, they simply failed to notice some relevant facts about what is intuitive.

**SECTION IV: Conclusion**

There is no question that our intuitions about hypothetical cases display order effects. The only question is about whether such order effects can be used to criticize the hypothesis of Intuitional Reliability. I have argued that they cannot. My argument, to recap, goes like this:

(1) It is intuitive that S knows that p only if S has no substantial defeaters of the reliability of the cognitive competence that S exercises in believing that p.

(2) It is intuitive that the notion of *substantial defeat* is vague, and that a defeater can be more or less substantial.

(3) It is intuitive that the Truetemp case involves a borderline case of substantial defeat (Truetemp has a defeater of the reliability of his competence, but it's not a very substantial defeater).

(4) It is intuitive that, for a psychologically salient property X (e.g., how substantial a defeater is), when a borderline case of X is compared against clear cases of X, it seems less X-ish, and when a borderline case of X is compared against clear cases of non-X, it seems more X-ish.

(5) What is intuitive is, in normal circumstances, by and large true. (Intuitional Reliability)

(6) So, 1, 2, 3, and 4 are all probably true. (From 5)

(7) Therefore (probably), when the Truetemp case is compared against a clear case of substantial defeat (like Coinflip Dave), it seems less like a case of substantial defeat (and so more like a case of knowledge), and it is probably true that when the Truetemp case is compared against a clear case of no substantial defeat (like Karen the Chemist) it seems more like a case of substantial defeat (and so less like a case of knowledge).

The hypothesis of Intuitional Reliability implies that it is generally reasonable to accept – independently of any corroboration they may receive from other sources – the verdicts rendered by our own intuitions concerning whether a particular property is instantiated in a particular hypothetical case. If this hypothesis is true, then this supports the conclusion that it is reasonable to accept – independently of any corroboration they may receive from other sources – the verdicts rendered by intuition concerning the intuitiveness of describing the Truetemp Charles case as a case of knowledge when it is contrasted with one or another different case. It also supports the conclusion that it is reasonable to accept – independently of any corroboration they may receive from other sources – the verdicts rendered by our own intuitions in favor of the no substantial defeaters hypothesis.

In other words, Intuitional Reliability predicts that we will find the very order effects that SAW found, and it also predicts the truth of a principle that explains why we find these very order effects. If Intuitional Reliability can predict and explain all of SAW’s data, then that seems to me to be some considerable evidence in favor of Intuitional Reliability.[[5]](#footnote-5)

1. *Philosophy and Phenomenological Research* 76(2008), pp. 138-9. [↑](#footnote-ref-1)
2. Though see Jen Wright, “On Intuitional Stability: The clear, the strong, and the paradigmatic” (*Cognition* 115 (2010):491-503), for more data bearing on SAW’s hypothesis. [↑](#footnote-ref-2)
3. “Experimental Philosophy and Philosophical Intuition,” *Philosophical Studies* 132 (2007): 99-107. [↑](#footnote-ref-3)
4. Cf. Weinberg, Jonathan M., Shaun Nichols & Stephen Stich. 2001. “Normativity and Epistemic Intuitions.” Philosophical Topics 29: 429-460. [↑](#footnote-ref-4)
5. Thanks to James Beebe and Jonathan Weinberg for helpful comments on an earlier draft of this paper. [↑](#footnote-ref-5)