

# **Testing for the phenomenal: Intuition, metacognition, and philosophical methodology**

Miguel Egler, Arché Research Centre, University of St Andrews, UK

**Forthcoming in Mind & Language**

**Accepted: 30<sup>th</sup> of August, 2018**

## **Abstract**

Recent empirical studies raise significant methodological concerns about the use of intuitions in philosophy. According to one prominent line of reply, these concerns are unwarranted since the empirical studies motivating them do not control for the putatively *characteristic* phenomenology of intuitions. This paper makes use of research on metacognitive states that share the phenomenology of intuitions to argue that this reply fails. Furthermore, it shows how empirical findings about these metacognitive states can help philosophers make better informed assessments of their warrant for relying on intuitions in inquiry.

## **KEYWORDS**

intuition, metacognition, phenomenalism, philosophical methodology, experimental philosophy, feeling of rightness

## **1. INTRODUCTION**

Recent empirical studies suggest that people’s intuitions show marked variation with respect to a host of epistemically irrelevant factors, such as their culture, or the order in which experimental tasks are presented to them.<sup>1</sup> A common reading of these findings is that they raise significant methodological concerns about the use of intuitions in philosophy. I call this the “Experimentalist challenge”. In response, some have argued that empirical studies about

---

<sup>1</sup> For a review, see Machery (2017, ch. 2).

intuitions are in need of considerable refinement before they can be put to use in debates about the methodology of philosophy. The most promising articulation of this proposal builds on the thesis known as Phenomenalism—that is, the view that intuitions are mental states defined by way of their *characteristic* phenomenology. Phenomenalists contend that since empirical studies about intuitions do not account for this phenomenology, it is unclear whether their findings actually reflect evidence of the workings of intuitions. Thus, Phenomenalists argue that we lack sufficient empirical evidence to conclude that intuitions vary in problematic ways, and that the Experimentalist challenge is thereby unmotivated. I dub this the “Phenomenalist rejoinder”.

This paper makes innovative use of research on human metacognition to shed light on debates surrounding the Experimentalist challenge. In more detail, it argues that a group of metacognitive states possesses the features that Phenomenalists take to be characteristic of intuitions. It then builds on this proposal to argue that the Phenomenalist rejoinder fails, and offers a set of methodological suggestions that help philosophers make better use of intuitions for the purposes of philosophical inquiry.

Here is a detailed plan of the paper: I first explain and distinguish different versions of the Experimentalist challenge (Section 2), and rehearse the central arguments for the Phenomenalist rejoinder (Section 3). I then briefly expound recent developments from research on human metacognition and argue that a particular set of *metacognitive states*—namely, a group of mental states with a high feeling of rightness—can be aptly characterized as intuitions (Section 4). As we will see, this proposal provides good reasons to think that the Phenomenalist rejoinder fails. In the last section (Section 5), I argue that research on human metacognition pinpoints sources of error in intuitions, and can inform philosophers of when they ought to refrain from relying on them in inquiry.

## 2. THE EXPERIMENTALIST CHALLENGE

Intuitions play a central role in philosophy, often shaping debates about prominent philosophical controversies. The intuitions in question concern people’s verdicts about whether or not some philosophical notion applies to a given (hypothetical or actual) scenario. Philosophers invoke these intuitive verdicts in philosophical arguments—generally in accordance with the principle that theories consistent with widely shared intuitions should be preferred to those that are not.<sup>2</sup> One clear example of this approach are the recent debates about the “Trolley problem”—that is, hypothetical scenarios where people are asked if it would be morally permissible to alter the

---

<sup>2</sup> See, e.g., Pust (2000).

course of a run-away trolley to save many people at the cost of sacrificing the life of one person (Thomson, 1985). Typically, theories about the moral permissibility of actions are required to accommodate for the fact that many people judge it morally permissible to change the course of the trolley, under the premise that this intuitive verdict is somehow representative of moral truth.<sup>3</sup>

The Experimentalist challenge raises doubts about these intuition-based methodologies of philosophy. This challenge is driven by evidence that intuitions are sensitive to a host of factors which have no impact whatsoever on philosophical truths.<sup>4</sup> For example, in a study about the Trolley problem (explained above), Petrinovich and O’Neill (1996) report that people’s intuitions about this case show marked variation with respect to superficial changes in how this problem was *worded*. Specifically, they found that people were less likely to judge it morally permissible to change the trolley’s direction when they read “Throw the switch, which will result in the death of the one innocent person on the side track”, when compared to people who read “Throw the switch, which will result in the five innocent people on the main track being saved”. These findings, along with other similar evidence of worrying patterns of variations on intuitions, putatively motivate at least *some degree of scepticism* about the use of intuitions in philosophy.<sup>5</sup>

Now, it is still a matter of heated debate just what degree of scepticism is warranted by the above body of findings. On more *radical* proposals, these findings are taken to show that intuitions provide *no warrant whatsoever* to philosophical theories (Stich, 2009; Swain *et al.*, 2008). However, proponents of *moderate* articulations of the Experimentalist challenge disagree: They claim instead that appeals to intuitions for philosophical purposes are *sometimes* warranted, but that they should be accompanied by a better understanding of when they are prone to leading us astray so as to avoid appealing to them in such cases (Weinberg, 2015).

In what follows, I explore the viability of the Phenomenalist rejoinder to the Experimentalist challenge. According to this line of reply, empirical studies about intuitions fail to motivate any degree of scepticism whatsoever about their use in philosophy. To begin examining this proposal, in the next section I briefly rehearse the main arguments for Phenomenalism and detail how its proponents elaborate on this thesis to advance the above rejoinder to the Experimentalist challenge. A careful and detailed exposition of these arguments will prove important to articulate two central claims of this paper. The first is that a particular class of metacognitive states—namely, a group of mental states that have a high *feeling of rightness*—share this putatively characteristic phenomenology of intuitions. The second is that empirical evidence of the workings of these metacognitive states undermine the Phenomenalist

---

<sup>3</sup> For a discussion, see Kamm (2015).

<sup>4</sup> For a review: Machery (2017, ch. 2).

<sup>5</sup> See, e.g., Alexander and Weinberg (2007).

rejoinder to the Experimentalist challenge.

### 3. PHENOMENALISM

#### 3.1 The phenomenology of intuitions

The linchpin of all recent Phenomenalist accounts is the idea that intuitions and perceptual experiences are fundamentally similar kinds of mental states. A first line of support for this idea is that ordinary language seems to suggest as much: For example, we often speak as if we can come to *see* by intuition that there cannot be four-sided triangles, or that intuitions make it *seem* to us that killing a person to save five others is immoral. Phenomenalists take these metaphors as illustrative that *what it feels like* to have an intuition is similar to *what it feels like* to have a perceptual experience (Bengson, 2015; Chudnoff, 2013; Koksvik, 2011). In more detail, Phenomenalists suggest that both intuitions and perceptual experiences possess a (broadly defined) *presentational phenomenology*.

Examples help to clarify what this “presentational phenomenology” of perceptual experiences amounts to. For instance, consider the difference between a case where a subject S *sees* that a wall is red and a case where she is *told* that it is red. Phenomenalists propose that only S’s perceptual experience will have a phenomenology that makes it seem that the wall is red and that presents S with the *redness* of the wall. In more detail, they contend that only S’s perceptual experience will make it seem to her as if she is aware of a relevant property of the wall (viz., its phenomenal redness) that itself motivates her to endorse that the wall is red. Furthermore, they suggest that intuitions also possess a presentational phenomenology. Thus, a subject S’s intuition that *p* will both (i) make it seem to S as if *p* is true, and (ii) *present* S with relevant properties of *p* that motivate her to endorse *p* (Bengson, 2015; Chudnoff, 2013; Koksvik, 2011). Again, examples are helpful to clarify. Consider the following two propositions:

**Proposition 1** Two non-concentric circles have at most two common points.

**Proposition 2** If a quadrilateral is inscribed in a circle, the sum of the products of the two pairs of opposite sides is equal to the product of the diagonals

Chudnoff expects that many people will have an intuition that Proposition 1 is true after they consider this claim (Chudnoff, 2013, pp. 50–51). In more detail, he predicts that by imagining and manipulating the mental imagery of two circles, it will (i) seem to readers as if Proposition 1 is true, and (ii) that they will be presented with the properties of circles that seem to make it true

that they intersect in at most two points—which motivates the reader to endorse Proposition 1. By contrast, he expects that most people will not have an intuition about Proposition 2 when they consider this claim. That is, he predicts that thinking about a quadrilateral inscribed in a circle is unlikely to elicit a mental state in readers which both makes it seem to them as if Proposition 2 is true, and that presents them with the relevant properties of this figure that motivate them to endorse Proposition 2.

By comparing and contrasting the phenomenology of intuitions with that of other mental states, Phenomenalists seek to establish the following two claims:

**Presentational thesis (PT)** All intuitions possess a presentational phenomenology.

**Irreducibility thesis (IT)** Intuitions are irreducible to any other mental state.

Phenomenalists claim that these examinations lend support to PT insofar as they reveal that intuitions within a host of different domains (mathematics, logic, and philosophy) share a same presentational phenomenal character. Furthermore, they argue that these examinations suggest IT as they reveal that other mental states lack this same phenomenology. More precisely, Phenomenalists argue that there are four features of the phenomenology of intuitions that discern them from any other mental state (Bengson, 2015, pp. 720–725; Chudnoff, 2013 Chs. 1,2; Koksvik, 2011, Ch. 5). I detail these below.

First, intuitions are *conscious, non-perceptual propositional attitudes* that incline assent to what they represent. In other words, they are intentional states that do not involve any of the sensory modalities of perception, and that represent a proposition as true—thus motivating endorsement of that proposition. Second, they are *spontaneous* in that we do not have to weigh considerations for or against a given proposition before having an intuition about it. For instance, we do not deliberate on whether circles do, or do not intersect in at most two points before having an intuition that this is the case. Rather, it just simply strikes us that this is so once we manipulate the mental imagery of two circles. Third, intuitions are *autonomous*, insofar as they are not responsive to contrary evidence. That is, it is possible for a subject S to have an intuition that *p*, even if she reflectively endorses, believes, or judges that *not-p*. To illustrate this point, Phenomenalists note that many people report having a persistent intuition that the naïve comprehension axiom<sup>6</sup> is true, even though they *know* that it is false. And lastly, intuitions are *seemingly truthful*, in that they purport to make subjects aware of facts. In this sense, Chudnoff suggests that the reader’s intuition that Proposition 1 is true will seem to make them aware of

---

6 The naïve comprehension axiom states that for every arbitrary condition *x*, there will be a set containing all, and only the things meeting that condition *x*.

*why* it is true. In more detail, he proposes that this intuition will enable thoughts about the properties of how circles intersect, which seem to make it true that they have at most two common points (Chudnoff, 2013, pp. 50–51). On a similar note, Bengson suggests that intuitions make subjects aware of features of their contents that seem to rationalize assent, “in the (psychological) sense that they tend to make formation of corresponding beliefs seem rational or fitting from the first-person perspective” (Bengson, 2015, p. 723). Thus, he proposes that intuitions make it seem “from the inside” as if their contents are “worthy of belief”. Likewise, Koksvik suggests that, part of what it is like to have an intuition is that it “purports to represent an objective fact” (Koksvik, 2011, pp. 168).

Now, Phenomenalists do not intend for these four features to give an exhaustive account of the phenomenology of intuitions. Nevertheless, they take them as *sufficient* to demarcate intuitions from any other kind of mental state (Bengson, 2015; Chudnoff, 2013, Ch. 1; Koksvik, 2011). For instance, they suggest that intuitions are markedly distinct from doxastic states (e.g., beliefs and judgments), insofar as the latter are *not* autonomous—since we believe, judge, or are inclined to believe or judge in accordance with the evidence available to us. They also argue that intuitions are unlike mere guesses or hunches, which they contend are *not* seemingly truthful—as they *at most* incline one to believe or judge, without making their contents seem worthy of belief. And so on and so forth for all other kinds of mental states, none of which, they propose, possess all the above four features.

Critically, for the purposes of this paper I will grant both PT and IT. That is, I take it that a presentational phenomenology is a marker of intuitions, and that an analysis in terms of the above four features suffices to distinguish intuitions from any other mental states. Accordingly, I henceforth use the unqualified term “intuition” to refer to mental states with a presentational phenomenology. In this way, I aim to give the most charitable interpretation of the Phenomenalist views and of their reply the Experimentalist challenge, which I now turn to.

### **3.2 A rejoinder to the Experimentalist challenge**

The proposal that intuitions are defined by way of their phenomenology motivates a forceful critique of the Experimentalist challenge. Central to this reply is a dispute over a common assumption operative in empirical studies motivating this challenge: Namely, the proposal that people’s answers on questionnaire-style surveys amount to reports of intuitions. However, if the Phenomenalist thesis is warranted, it suggests that this assumption is questionable. After all, mere reports of a person’s answers are not instructive as to whether they had an experience with a presentational phenomenology when responding to that task. Bengson (2013) provides a very clear formulation of this proposal. Commenting on the studies motivating such methodological concerns, he suggests that:

[T]hese attacks neglect a considerable gap between the answers elicited by the relevant empirical studies and the intuitions about which naysayers naysay. It cannot innocently be assumed that subjects' answers expressed how things struck them—what intuitions they had, if any. The point is simple, but not insignificant. For, I will argue, it implies that we are at the present time unwarranted in drawing any negative conclusions about intuitions from the relevant empirical studies. (Bengson, 2013, p. 496)

To buttress this proposal, Phenomenalists examine prominent studies in experimental philosophy with a view to teasing out how their findings might conflate intuitions with other mental states—for example, hypotheses, guesses, emotional reactions, inferences, or other conclusions (Bengson, 2013, p. 497; Chudnoff, 2013, pp. 107–113). For instance, Bengson contends that the studies showing *wording effects* on the Trolley problem might be running together intuitions with mere emotional reactions. Specifically, he contends that because such studies highlight the negative outcome of the scenario (i.e., that one person will be killed), they trigger negative emotions, which leads people to refrain from opting to change the course of the trolley (Bengson, 2013, pp. 517–518). Building on such examinations, Phenomenalists suggest that there are decisive reasons to think that empirical studies about intuitions are simply inadequate to motivate the Experimentalist challenge.

In the next section I argue that there are good reasons to think that this Phenomenalist rejoinder fails. As we will see, recent inquiry into human metacognition has studied a class of psychological states that share crucial phenomenological commonalities with intuitions. I then discuss evidence that these psychological states display strikingly similar patterns of variation as those uncovered by studies in experimental philosophy. Lastly, I explain why these findings undercut the Phenomenalist rejoinder to the Experimentalist challenge. To set up this argument, I start out in the next section by summarizing key features of recent research on human metacognition.

#### **4. METACOGNITION AND PHENOMENALISM**

Imagine someone asks you “Who wrote *A Clockwork Orange*?”. You stop for a moment and try to recall; you are sure that you know the answer, but the name just escapes you now. Or suppose you are given a piece of paper with a couple of math problems. You try to solve the first one, but you feel as if you have got it wrong. You move on to the next, which looks much easier. You then become pretty confident that you got it right. Common to all these events is the operation of human capacities of metacognition: The set of capacities that allow for representation and

monitoring of one's own cognition (Proust, 2013).

Much of the recent research on metacognition has focused on the study of so-called *metacognitive experiences*—a number of phenomenal states that are *about* one's own cognition (Arango-Muñoz & Michaelian, 2014). For current purposes, I focus on the metacognitive experience known as the *feeling of rightness* (henceforth FOR; FORs for the plural “feelings of rightness”). This refers to a gradable impression that the answer arrived at in response to a cognitive task is correct. Low degrees of FORs yield a weak impression of accuracy and are correlated with higher rates of change in later responses. By contrast, high degrees of FORs give a strong impression that the answer produced is accurate, which make people less likely to change their responses later on (Thompson & Johnson, 2014; Thompson et al., 2011, 2013b).

Notably, there are at least very superficial similarities between mental states with a *high* FOR and intuitions: Namely, both have a phenomenology which makes their contents seem *true* to a subject. In what follows, I flesh out the similarities between intuition and mental states with FOR to defend the following claim:

**Int.FOR** Some mental states with a *high* FOR are intuitions.<sup>7</sup>

In support of Int.FOR, I first argue that a particular group of mental states with a high FOR display precisely the features that Phenomenalists take as sufficient to demarcate intuitions from other mental states (detailed in the previous section). I then detail why Int.FOR suggests that Phenomenalist rejoinder to the Experimentalist challenge fails.

#### 4.1 Phenomenalism and high FOR

Investigations into the workings of the FOR make use of a number of reasoning tasks to elicit and examine this metacognitive experience (Thompson et al., 2011, 2013b). These experimental studies take on a very similar approach to the one that Phenomenalists use to argue for their views: Namely, they invite readers to entertain particular examples, and to then immediately attend to the mental states that arise. To illustrate the type of task used in research on the FOR, consider the following pair of examples:

**Ex. 1** If a car runs out of gas, it will stall. The car has not stalled, therefore it has not run out of gas.

---

<sup>7</sup> This thesis is not to be read as a bi-conditional. Thus, I leave it open whether every intuition is a mental state with a high FOR.

**Ex. 2** None of the nurses are magicians. Some of the winemakers are nurses. Therefore, some of the magicians are not winemakers.

Alternatively, consider the following two examples:

**Ex. 3** Suppose you are told that a particular pub follows the rule that if a person drinks alcohol, then they must be at least 18 years old. Of the four people at the pub, the first is 16 years old, the second is 18 years old, the third is drinking beer, and the fourth is drinking water. To check whether the pub is actually enforcing their own rule, you need only ask what the 16-year old is drinking, and whether the person drinking beer is more than 18 years old.

**Ex. 4** Most puppies have a price. Some friendly animals have a price. Some puppies are friendly animals.

Research on the FOR suggests that examples like Ex. 1 and Ex. 3 are likely to elicit mental states with a high FOR in the reader. Thus, readers who consider Ex. 1 are expected to immediately accept that the car would indeed not have run out of gas, and for it to seem to them as if it is *true* that this conclusion follows from the premise. Similarly, reading Ex. 3 is predicted to make it seem true that one need only ask the 16-year old what they are drinking, and how old is the person drinking beer, to check whether the pub is enforcing the rule. Thus, evaluating Ex. 1 and Ex. 3 will elicit experiences that are categorically distinct from a mental state with a low FOR such as a mere *guess* or *hunch*, which can at most incline accepting or rejecting a statement, but which does not make it seem true. Conversely, Ex. 2 and Ex. 4 are unlikely to elicit a mental state with a high FOR: Most readers who entertain Ex. 2 and Ex. 4 will not immediately arrive at an answer that accompanies a feeling as if this response is accurate.

These brief illustrations underscore superficial similarities between intuitions (as these are defined by Phenomenalists) and mental states with a high FOR: Namely, both make their contents seem true. However, further careful reflection on these examples of high FORs brings to light some other very significant and important commonalities with intuitions. Most notably, these mental states with a high FOR have a phenomenology that is very similar in kind to that which Phenomenalists ascribe to intuitions. For instance, I expect that considering Ex. 1 will both make it seem to readers as if this claim is true and that they will be presented with that which makes it true. That is, this mental state will make it seem to readers as if they are *aware of properties of* Ex. 1 that motivate endorsing this claim—for example, that Ex. 1 is just a case of *modus tollens*, which makes it true that the car would not have run out of gas. Likewise, I expect

that entertaining Ex. 3 will make this claim seem true to the reader and will present them with properties about this claim that motivate endorsing Ex. 3. In this sense, it will seem clear to readers *why* one need only ask what the 16-year old is drinking and the age of the person drinking beer to check whether the bar is following its own rule.

Furthermore, there is additional support for the parallel between intuitions and the above examples of mental states with a high FOR in that the latter are also aptly characterized in terms of the four features that Phenomenalists take as sufficient to demarcate intuitions from all other mental states. First, the mental states with high FOR above are conscious non-perceptual propositional attitudes that incline assent to what they represent. That is, the mental states that ensue from considering both Ex. 1 and Ex. 3 are intentional states that represent a proposition as true, thus motivating endorsement of that proposition, but which do not involve the sensory modalities of perception. Second, these examples of mental states with a high FOR are also *spontaneous*: They simply occur without the need to weigh considerations for or against a certain proposition. Thus, I expect that readers who consider Ex. 1 will not have to deliberate about any one position or other about whether a car has run out of gas before having a mental state with a high FOR that Ex. 1 is correct. Rather, it will simply strike readers as true that the car has not run out of gas. Third, these mental states with a high FOR are *autonomous* in the sense that they do not respond to contrary evidence. Indeed, a large body of empirical research suggests that people are prone to have mental states with a high FOR about claims which they do not judge or believe to be true (Ackerman & Zalmanov, 2012; Thompson et al., 2013a, 2013b).

And lastly, these mental states with a high FOR are *seemingly truthful* in that they purport to make subjects aware of facts. In line with Chudnoff's explanation of what this amounts to, I propose that a mental state with a high FOR about Ex. 1 will seem to make one aware of *why* this proposition is true. For instance, I expect that when the reader considers Ex. 1, it will seem to them as if they are made aware of the precise logical properties instantiated by Ex. 1 which make it true—for example, readers will recognize that Ex. 1 is a case of modus tollens. Furthermore, in line with Bengson's proposal, I submit that identifying Ex. 1 as a case of modus tollens will make it seem rational from the reader's perspective to form a corresponding belief that Ex. 1 is true. On a similar note, I predict that when the reader considers Ex. 3, it will seem to them as if they are made aware of *why* this claim true. In this sense, I expect that the reader will be made aware of the logical relations that make it true that one need only ask two of the people at the bar to check whether the pub is following its own rule, and that this will make a belief in Ex. 3 seem rationally fitting from the reader's perspective.

This additional analysis suggests that the mental states with a high FOR that ensue from considering Ex. 1 and Ex. 3 share a set of important features with intuitions. Specifically, like intuitions, they are also aptly described as being *conscious non-perceptual* propositional attitudes that incline assent to their contents, and which are *spontaneous, autonomous, and seemingly*

*truthful* in character. Moreover, to the extent that intuitions are characterized by way of these features, this analysis suggests that the mental states with a high FOR elicited above can be aptly described as intuitions—that is, they lend support to the thesis I have called Int.FOR. So, in what follows, I consider the mental states with a high FOR that arise from considering Ex. 1 and Ex. 3 to be intuitions.

For the remainder of this section, I assume Int.FOR as a working hypothesis and proceed to tease out its implications. In particular, I argue that this thesis shows there to be substantial empirical support against the Phenomenalist rejoinder. To set up this argument, I will now detail evidence that mental states with a high FOR of a similar kind to those exemplified above display patterns of variation that are strikingly similar to those uncovered by studies in experimental philosophy.

## **4.2 Experimental findings**

A central ambition of studies into the workings of the FOR has been to identify the determinants of this metacognitive experience. Researchers have thus conducted studies with a view to examining whether people’s reports of FORs vary with respect to changes in certain defined parameters. Many of these studies report that mental states with a high FOR display patterns of variation that are surprisingly similar to those uncovered in studies from experimental philosophy. Specifically, they provide evidence of variation in people’s reports of high FORs depending on how questions and tasks were framed (Thompson et al., 2011), the order in which experimental tasks were presented to them (Markovits et al., 2015, exp. 2), as well as demographic factors, such as whether people were of high cognitive capacity or low cognitive capacity (Thompson and Johnson, 2014). I will now first briefly describe one such set of findings—namely, those showing that mental states with a high FOR very similar to Ex. 1 and Ex. 3 (discussed above) are subject to a variety of *framing effects*. I then build on these and other findings of worrying patterns of variation in mental states with a high FOR to argue that the Phenomenalist rejoinder to the Experimentalist challenge fails.

Thompson and colleagues (2013a) report evidence of framing effects on mental states with a high FOR on the famous Wason Selection Task (henceforth, WST). We have already seen an example of this task. That is because Ex. 3 is one version of the WST. On more traditional formulations of the WST, people are presented with pictures of four cards and informed that each card has a letter printed on one side, and a number on the other. Two of the cards display the side printed with a number, and the other two display the side with the letter (e.g., the cards could be labeled as “A”, “S”, “3”, and “7”). People are then asked to read a conditional statement expressing a rule about the relation between the number and the letter of a specific card (e.g., “If a card has the letter ‘A’ on one side, it has the number ‘3’ on the other”). Then, they are asked to identify which of the cards shown need to be turned so as to verify whether the set of cards being

shown are compatible with this rule—that is, if they do not violate this rule.

Most people give the wrong response on this traditional—so-called “implicit negation”—version of the WST. For instance, when asked which cards would verify the rule “If a card has the letter ‘A’ on one side, it has the number ‘3’ on the other”, people either choose to flip the cards showing the “A” and the “3”, or just the card displaying the “A”. (The correct answer is in fact to select the cards labeled “A” and “7”.) However, people are much less prone to giving this wrong answer on an explicit negation version of the WST. On this version of the task, all the cards are labeled with either the letter or number mentioned in the rule or their negation—such that the rule “If a card has a letter ‘A’ on one side, it will have a ‘3’ on the other” would be followed by cards labeled “A”, “not-A”, “3”, or “not-3” (Evans et al., 1996; Stahl et al., 2008). And similarly, people also tend to give the right answer to the WST, when this task is formulated in terms of everyday social situations, such as in Ex. 3.

Thompson and colleagues (2013a) conducted an experiment aiming to further examine this shift in people’s responses to the WST. In this study, they asked people to complete either an implicit or explicit negation version of the WST within a novel two-response paradigm used to study the FOR. They asked people to first issue a quick answer to the WST and evaluate their degree of confidence about this response; in a second experimental phase, they asked people to reconsider their first answer in free-time and to evaluate their confidence in this second response. In line with previous findings, Thompson et al. (2013a) also provide evidence that people in the implicit negation condition often gave the wrong answer to the WST. More importantly though, they found that these people tended to also attribute a high FOR about this answer and to endorse this answer when later asked to reconsider it. Conversely, people in the explicit negation condition were more likely to give correct responses to this task and to attribute high FORs to their answers—also often endorsing that answer when asked to reconsider. In effect, these findings thus reveal a framing effect on mental states with a high FOR about the WST. Specifically, they reveal that minor changes in the superficial features of the WST (viz., framing it either in terms of implicit or explicit negations) can determine the answers to which people report high FORs.

In another study, Thompson and colleagues (2011) provide additional evidence of framing effects in mental states with a high FOR. In this set of experiments, they asked people to evaluate what logically followed from a simple conditional statement of the form “If A then B”, again using the two-response paradigm described above. We have already seen one case of this kind of task above in Ex. 1. As in Ex. 1, all the conditional relations used in this set of experiments made reference to familiar objects and situations from everyday life; however, for half of the questions used, the conclusions were coded as being “unbelievable” in that they were likely to contradict people’s background knowledge about the objects and situations there mentioned. For instance, the following is a case of such an unbelievable conclusion: “If the TV is

plugged in, then it works; the TV is plugged in, therefore, it works”. Most people know that TVs can fail to work even if plugged in (they may be faulty, or power may be out). Knowledge of this fact makes the above conclusion strike most people as implausible (“unbelievable”), even if that conclusion does follow logically from the premises. The remaining conclusions in the experiment were framed so as to be made believable in that they were compatible with background beliefs about the items mentioned in the task—for example, “If a car runs out of gas, then it will stall; if the car has not stalled, therefore it did not run out of gas”.

Thompson et al. (2011) found that participants were more likely to evaluate that believable conclusions followed from the premises and that they frequently attributed high FORs about this response, even when the inference was invalid. However, people often judged unbelievable conclusions to *not* follow from the premises and to also report high FORs about these responses, even when the inference was valid (for similar findings, Thompson et al., 2013b). In effect, this pattern of evaluations pinpoints a framing effect in mental states with a high FOR. More precisely, they show that framing conclusions as either believable or unbelievable can have a significant impact on whether people have a high FOR about whether it follows or not from a given set of premises.

Besides these two sets of findings, additional studies have uncovered similarly worrying framing effects on mental states with high FORs in evaluations of conditional statements—similar to that detailed in Ex. 1 (Ackerman and Thompson, 2017). Furthermore, inquiry into the FOR has also found other worrying patterns of variation in mental states with a high FOR to similar tasks among people with high or low cognitive capacity (Thompson & Johnson, 2014), and the order in which cases are presented to people (Markovits et al., 2015, 2017).

I contend that, when taken together, these studies raise significant doubts about the Phenomenalist rejoinder. Recall, this amounts to the suggestion that the Experimentalist challenge is ultimately unmotivated, insofar as we lack sufficient evidence to conclude that intuitions—rather than a host of other mental states—vary with respect to the parameters identified in studies in experimental philosophy. But, I have argued that the mental states with high FOR that arise from considering Ex. 1 and Ex. 3 can be aptly defined as intuitions. Furthermore, I have argued that the empirical literature shows that the mental states with high FOR like Ex. 1 and Ex. 3 vary with respect to similar kinds of factors to those used to motivate the Experimentalist challenge.<sup>8</sup> In light of this, I submit that there is mounting evidence that the Phenomenalist rejoinder fails.

### 4.3 Objections

I now turn to a brief exposition and rejection of what I take to be the two most promising replies

---

<sup>8</sup> For additional empirical support, see also Danek and Wiley (2017), and Topolinski and Reber (2010).

available to Phenomenalists against the arguments in this section. The first is to claim that findings about mental states with a high FOR discussed above are inadequate to assess the Phenomenalist rejoinder. This reply builds on the observation that the examples used to elicit these mental states with a high FOR involve only simple conditional statements or arguments. However, these examples are not the kinds of cases that are typically used in the philosophical literature to elicit people's intuitions—for example, Gettier scenarios, or the Trolley problem. And since the Phenomenalist rejoinder refers to the use of intuitions in *philosophy*, then there are reasons to doubt that the findings discussed above actually concern the intuitions that are at stake in the Experimentalist challenge. As such, Phenomenalists might claim that there is little reason to think that those findings show that *philosophical* intuitions are prone to vary in problematic ways.

Although seemingly plausible, there are significant difficulties for this line of reply. The central issue is that it relies on tracing a significant distinction between philosophical and non-philosophical intuitions. However, Phenomenalists maintain that all intuitions—regardless of their content (e.g., mathematical, logical, philosophical)—are categorized by way of their shared presentational phenomenology, which demarcates them from any other mental state. On this view, there are no intrinsic and fundamental differences among intuitions at the level of their subject matter that would warrant claims to the effect that philosophical intuitions are more reliable than any other. As such, I propose that the only option available to Phenomenalists to salvage this first line of reply would be to claim that the mental states with a high FOR I have examined are not in fact intuitions. I take it that the most promising way to flesh out this proposal is to argue that, since the examples used to elicit those mental states with high FOR involve evaluations of the validity of conditional statements, then they are actually *inferences*—not intuitions. This then raises doubts about the attempt to make use of findings about FORs to put pressure on the Phenomenalist rejoinder.

However, this alternative way of articulating the first reply fails by the Phenomenalists' own lights. That is because Phenomenalists themselves have advanced arguments to the effect that considering conditional statements can give rise to an intuition about their validity—such that it will seem to one that it is valid or not, and make it seem that one is *presented* with that which makes it so (Chudnoff, 2013, pp. 149–150; Koksvik, 2011, p. 177). For example, Chudnoff suggests that considering the following simple argument:

- (1) Every even number is divisible by two.
- (2) The number of pigs in the pen is even.
- (3) So, the number of pigs in the pen is divisible by two.

can give rise to “an intuition experience that represents that (1) and (2) support (3)”, which will seem to *present* one with why this is the case (Chudnoff, 2013, p. 150). Likewise, Koksvik (2011, p. 177) contends that considering the claim, “If my shoes are by the door, then they are not not by the door” will give rise to an intuition that *presents* this inference as valid. Similarly, I propose that the kinds of mental states with a high FOR examined in this paper are naturally described as intuitions of just this sort: That is, they are intuitions about the logical validity of an inference. After all, as argued for above (Section 4.1), these mental states with a high FOR do possess the characteristic (presentational) phenomenal character of intuitions, and are thus aptly defined as intuitions. Given these claims, we then return to the first point: If such intuitions about the logical validity of conditional statements are supposed to be markedly distinct from philosophical intuitions—for example, such that only the former are prone to the worrying patterns of variation found above—then Phenomenalists would owe us an explanation for why this should be the case. Even so, I should note that it is quite difficult to see just why intuitions about the logical validity of inferences are in any way philosophically irrelevant. For example, consider the kinds of informal arguments that philosophers routinely advance to the effect that a given theory implies a rather unpalatable conclusion, which is taken to suggest that the theory is thereby false. In such cases, it is very reasonable to expect that philosophers often rely on intuitions about what logically follows from the central claims of that theory to evaluate whether that theory does indeed imply that problematic consequence. In this sense, I take it that intuitions about the logical validity of inferences are relevant for debates surrounding the Experimentalist challenge (see Section 5 for additional discussion of this point). For these reasons, it is difficult to argue that the empirical findings about mental states with high FOR discussed above are inadequate to evaluate the Phenomenalist rejoinder.

A second promising line of reply to the arguments in this section is an objection analogous to the Phenomenalist rejoinder, yet aimed at studies on the FOR. The main thrust of this type of reply is the idea that findings from these experimental studies are likely to conflate mental states with high FORs with a variety of other mental states, such as people’s post-hoc rationalizations, beliefs, guesses, and hunches. Building on this proposal, Phenomenalists might object to the suggestion that findings from inquiry into the FOR are apt to motivate methodological concerns about the use of intuitions in philosophy.

Although seemingly plausible, a main difficulty for this objection is that empirical studies on the FOR have made considerable efforts to avoid conflating mental states with a high FOR with other such confounding phenomena. In particular, they make use of a novel two-response experimental paradigm that has been found effective for this end (Thompson et al., 2011). On the first experimental stage of this framework, people are asked to quickly give an answer to reasoning problems and to then immediately rate their FOR about this initial answer (marking this on a scale from one to seven). Adoption of this first quick-response paradigm is motivated by

empirical findings which show that asking people to give their first immediate judgments about a task interferes with their ability to engage in deliberate reflection about the experimental task at hand (Neys, 2006), to prompt less neural activation from areas associated with belief inhibition (Tsuji & Watanabe, 2010), and to lead people to report those answers that immediately strike them as plausible—rather than what they infer or suppose is the correct response to that task (Evans & Curtis-Holmes, 2005). This first stage of the experimental design thus helps researchers home in on mental states with high FORs by minimizing the possibility that people’s responses reflect their background beliefs, inferences, and suppositions about that task.

The second-stage of the two-response paradigm used in inquiry into FOR is similarly well motivated. On this second stage, people are asked to reconsider their initial response to the experimental task—now in free-time. A robust finding from studies on FOR is that people’s reports of a high FOR about their answer to an experimental task is a very good indicator that they will endorse that answer when asked to reconsider it.<sup>9</sup> Owing to this, mental states with high FORs are regarded as distinct from a simple hasty or unreflective response—of the sort one would quickly reject upon closer scrutiny. This indicates that the second stage slow-response paradigm is also suitable to inquire into the FOR. In particular, to the extent that this allows researchers to track changes in people’s first and second answers, and to correlate them with reports of high FORs, this helps to distinguish between mental states with high FORs from other confounding phenomena—such as mere hasty or unreflective responses.

In addition, studies on the FOR have also actively sought to rule out the possibility that their findings might reflect people’s mere guesses or hunches about the experimental task (Thompson et al., 2011, 2013a). As previously mentioned, both of these are defined as mental states with a low FOR, insofar as they at most incline subjects to assent to a claim without making it seem as if that claim is true. In order to rule out interference of guesses and hunches in their findings, researchers instructed participants in their experiments to report FORs on a Likert scale ranging from one to seven, in which the lower ends of this scale were clearly labeled as a mental state with a low FOR—for example, “Just guessing”. In effect this means that people’s reports of high degrees of FOR in these experiments were made in explicit contrast to self-reports of mere guesses or hunches.

Taken together, these considerations raise significant difficulties for the claim that studies about the FOR might conflate mental states with a high FOR with a variety of other phenomena. As detailed above, the two-stage experimental paradigm used in inquiry into FOR has been shown effective at minimizing the possibility that experimental findings might reflect evidence of participants’ post-hoc rationalizations, beliefs, judgments, inferences, and mere guesses or hunches. As such, the concerns that Phenomenalists have raised about studies in experimental

---

<sup>9</sup> For a review of the evidence, see Ackerman and Thompson (2017)

philosophy do not naturally carry over to studies on the FOR.

#### **4.4 Summing up**

I have been arguing that some mental states with a high FOR can be aptly defined as intuitions (Int.FOR), and that these mental states are prone to patterns of variation that are strikingly similar to those uncovered in studies in experimental philosophy (e.g., framing effects, the order in which cases are presented, and demographic variables). I then proposed that these findings indicate that intuitions do indeed vary in the ways that experimental philosophers have suggested, and that this refutes the Phenomenalist rejoinder to the Experimentalist challenge. Lastly, I discussed and rejected the most promising replies available to Phenomenalists against these arguments.

Now, one important upshot of these arguments is that they provide further support to the Experimentalist challenge. After all, they both undercut a prominent objection to these experimentally motivated methodological concerns, and give additional evidence that intuitions vary with respect to truth-irrelevant factors. However, as mentioned in the outset, it is still a matter of great dispute just what these findings imply about the use of intuitions in philosophy. Defenders of *radical* versions of the Experimentalist challenge propose that these findings show that philosophers should refrain from using them in inquiry. Proponents of the *moderate* Experimentalist challenge disagree: Rather, they suggest that appeals to intuitions in philosophy are sometimes warranted, but that these should be accompanied by a better understanding of when we can trust intuitions, and under what circumstances they are prone to leading us astray.

For the remainder of this paper I argue that Int.FOR lends support to a moderate Experimentalist challenge. To develop this proposal, I briefly describe the two most prominent formulations of *radical* versions of the Experimentalist challenge—namely, those proposing that intuitions are *unreliable* or *hopeless*—and argue that both of fail. As we will see, the arguments I advance against these views will also reveal how findings from inquiry into the FOR pinpoint sources of errors in intuitions, which inform philosophers how to make better use of them in inquiry.

### **5. ADVANCING THE DEBATE**

#### **5.1 Radical Experimentalist challenges**

A very common reading of findings from experimental philosophy is that they speak to the reliability of our intuitions. In this sense, evidence that particular intuitions vary with respect to truth-irrelevant factors are taken to show that those intuitions are unreliable. More controversially, many experimental philosophers regard these worrying findings as illustrative of

the epistemic deficiencies afflicting intuitions in general. Thus, they suggest that the charge of unreliability extends much wider than just these local demonstrations. Given the sensible assumption that we should refrain from using unreliable sources of evidence in philosophical inquiry, then the above considerations suggest that we should not make use of intuitions in philosophical inquiry.<sup>10</sup>

There are, however, good reasons to resist this strong conclusion. Specifically, a growing body of empirical work suggests that a number of philosophically relevant intuitions stem from cognitive processes that are in fact generally reliable. For instance, Jennifer Nagel (2012) has argued that intuitive knowledge attributions that are central to epistemology arise from the exercise of ordinary psychological capacities for “mind-reading”—that is, a set of operations humans routinely rely on to understand and predict the mental states of others. Given the considerable amount of evidence that capacities for mind-reading are very reliable—as shown by the accuracy of people’s predictions of others’ mental states—Nagel (2012) suggests that it is rather sensible to expect that intuitive knowledge attributions will be reliable too. A set of studies by Eugen Fischer and colleagues motivate a similar conclusion. In a number of experiments, they report evidence that some philosophically relevant intuitions are underwritten by processes that are key to our competence as speaker-hearers of a language—such as metaphor interpretation (Fischer, 2014) and use of stereotype-driven inferences in verb comprehension (Fischer & Engelhardt, 2016, 2017). Again, as there is substantial evidence that such processes are generally reliable—shown by how we successfully use them in everyday communication—then there are very good reasons to think that they give rise to reliable intuitions.

The central upshot of the above considerations is that they significantly weaken the claim that intuitions are generally unreliable. Thus, they undercut methodological concerns about the use of intuitions in philosophy that build on this proposal. However, it is noteworthy that even if we grant that intuitions are in fact generally reliable, this is still insufficient to show that philosophers are in any way warranted in appealing to them in inquiry. For instance, Jonathan Weinberg (2007) has argued that what is at issue in methodological concerns about the use of intuitions is whether they are “hopeless”. Weinberg suggests that a source of evidence provides warrant only if it is, in his technical sense, “hopeful”, by which he means that we are able to both identify and correct for that source’s errors when, and if, these arise (Weinberg, 2007, p. 327). Furthermore, he suggests that intuitions fare quite badly in this regard insofar as we simply lack an adequate understanding of when intuitions are prone to lead us astray, and what we can do to mitigate such errors. Moreover, although Weinberg does not endorse a radical Experimentalist challenge, these considerations can easily lend support to this view: For, given Weinberg’s claims that only “hopeful” evidential sources can provide warrant, and that intuitions are ultimately

---

<sup>10</sup> For a review, see Machery (2017, Ch. 3).

“hopeless”, then this clearly implies that intuitions are not apt to provide any warrant to philosophical positions and claims.

For the remainder of this paper, I argue that the claim according to which intuitions are ultimately “hopeless” fails to account for the many insights that research on the FOR provides into the workings of intuitions. As we will see in the next section, findings from this research helpfully single out particular factors that are prone to problematically influence people’s intuitions and can thus help philosophers identify and mitigate for such errors when they arise. Thus, the findings from inquiry into the FOR both undercut the above *radical* Experimentalist challenge and help to articulate a moderate version of this view.

## 5.2 A moderate Experimentalist challenge

I have previously discussed evidence that the perceived believability of an inference’s conclusion can influence people’s intuitions about logical validity (Section 4.2). In more detail, these findings show that people often judge inferences with “believable” conclusions to be valid, even when they are in fact invalid. Conversely, people are much more likely to judge inferences with “unbelievable” conclusions as invalid when they are in fact valid. One constructive way of reading these findings is that they helpfully pinpoint vitiating circumstances that can lead intuitions astray and which philosophers would do well to be wary of. Thus, they suggest a fairly simple set of practical recommendations that can potentially help improve intuition-based methodologies of philosophy: Take measures to minimize, or rule out, the possibility that people’s judgments of validity might reflect *merely* that which they find to be believable.

One way to lend weight to this proposal is to show how it proves apt to explain and help mitigate a worrying pattern of variation uncovered in a recent study by Nichols and Knobe (2007). In this study, they report evidence of a robust framing effect in people’s intuitions about moral responsibility and free will. They first presented people with the following description:

Imagine a universe (Universe A) in which everything that happens is completely caused by whatever happened before it. This is true from the very beginning of the universe, so what happened in the beginning of the universe caused what happened next, and so on right up until the present. For example, one day John decided to have French Fries at lunch. Like everything else, this decision was completely caused by what happened before it. So, if everything in this universe was exactly the same up until John made his decision, then it had to happen that John would decide to have French Fries. (Nichols & Knobe, 2007, p. 669)

People in the abstract condition were then asked “In Universe A, is a person fully responsible for their actions?”. Conversely, people in the concrete condition were asked to first read the description of Bill—an individual from Universe A who murders his wife and children and then runs away with his secretary—and were then asked “Is Bill fully morally responsible for his actions?”. Nichols and Knobe found that people answering the abstract question often judged that a person in Universe A would not be responsible for their actions, although people who read the case of Bill judged him morally responsible for murdering his family.

Evidence that the believability of conclusions might sway people’s intuitions about logical validity can aptly explain this difference in evaluations. To begin spelling this out, it is first helpful to get clear on the structure of the above experimental task. In particular, note that judgments to both the abstract and concrete questions are essentially evaluations of the validity of two variations of a logically identical conditional statement. Specifically, the abstract question asks readers to evaluate whether it follows that “If Universe A obtains, then any person in this universe is morally responsible for their actions”. Similarly, people in the concrete condition are in effect asked to evaluate whether it follows that “If Universe A obtains, then an individual in this universe called Bill (who murdered his family) is morally responsible for his actions”. We can describe the difference in evaluations in terms of the validity of inferences: Whereas people in the abstract condition judged the conditional statement as invalid, people in the concrete condition judged it to be valid.

Now, consider how the consequent of the conditional statement that mentions Bill is aptly described as being believable insofar as it focuses on the act of murder. After all, most of us strongly believe that murderers should be held responsible for their actions. Conversely, the conclusion of the more ‘abstract’ conditional can be described as unbelievable, as it focuses on no action in specific—and most of us do not believe that people should be held responsible for actions when they really had no other choice. Thus, the evidence that people tend to judge inferences with believable conclusions as valid and unbelievable conclusions as invalid seems apt to explain the above difference in evaluations. In this sense, people’s judgment that it follows that “If Universe A obtains, then an individual in this universe called Bill (who murdered his family) is morally responsible for his actions” can be attributed to the believability of this conclusion. Conversely, evaluations that it does not follow that “If Universe A obtains, then any person in this universe is morally responsible for their actions” is due to the fact that this conclusion is unbelievable. This then highlights that the proposed methodological recommendations are actually instructive. In particular, they illustrate a case in which philosophers would do well to attend to the potential deleterious effects of believable/unbelievable conclusions in intuitions about validity.

We can tease out some further helpful methodological suggestions from another set of findings from inquiry into the FOR. As previously discussed, Thompson and colleagues (2013a)

report evidence of framing effects on intuitions in the WST. In more detail, they found that people who complete the implicit negation version of this task—in which only two cards have lexical content matching those of the rule under examination—often choose to examine precisely the cards bearing superficial similarities to the rule. Conversely, people who complete the explicit negation version of this task—in which *all* cards have lexical content matching the rule—often choose a distinct set of cards. Thompson et al. (2013a) suggest that this finding illustrates how the similarity in the lexical content of questions and reasoning prompts can make particular elements appear more relevant to the task at hand, and thus induce people to focus on them when issuing answers. Again, we can read these findings as providing instructive methodological guidance for philosophers in that they point out that people’s intuitions about philosophical relevant scenarios might be swayed by lexical similarities between questions and the descriptions of cases. Thus, they suggest the following practical recommendation: Try to minimize, or rule out, the possibility that people’s intuitions might reflect *merely* their examinations of what is made more salient by the wording of the question.

To bolster this proposal, I show how it aptly explains a set of recent findings from a study by Alexander and colleagues (2018). In this study, they report evidence that people’s intuitions about many typical cases used in philosophical discussions about peer-disagreement display significant framing effects. For instance, in one experiment they presented people with the following variation of the very well-known “Restaurant Case”:

Suppose you and your friend go out to dinner. When it is time to pay the check, you agree to split the check evenly and to give a 20% tip. You do the math in your head and become highly confident that your shares are \$43 each. Meanwhile, your friend does that math in her head and becomes highly confident that your shares are \$45 each. You and your friend have a long history of eating out together and dividing the check in your heads, and know that you’ve been equally successful at making these kinds of calculations: usually you agree; but when you disagree, you know that your friend is right as often as you are. Moreover, you are both feeling sharp tonight and thought that the calculation was pretty straightforward before learning that you disagreed about the shares. (Alexander et al., 2018, p. 2540)

One group of people were then asked the question “Should you give your friend’s belief equal weight and think that it is no more likely that you are right than that your friend is right, or should you continue to prefer your own belief?”. Another group was asked instead, “How confident should you be that your belief is correct now that you know that your friend disagrees with you?”. Alexander and colleagues (2018) report that people in the first group often judged

that they should reduce confidence in their initial answer in this case, whereas people in the second group responded that they would maintain highly confident in it.

One plausible explanation for this finding is that the words used to formulate these questions highlight distinct elements of the Restaurant Case, thus prompting very different examinations of the scenario. To begin developing this proposal, note how the first question invites people to consider giving “equal weight” to a friend’s response, and asks whether their “friend is right”. In light of the evidence detailed above, it is plausible that this formulation highlights precisely the aspects of the Restaurant Case that are worded in a similar way: Thus, it brings attention to the description of the friend as “equally successful” and that “your friend is right as often as you are” in these kinds of mental calculations. In this sense, this particular framing of the question underscores the parts of the case that emphasize just how reliable the friend actually is; as such, it is quite unsurprising that people answering this first question often judged they should reduce confidence in their own belief in light of the disagreement with their friend.

Now contrast this with the formulation of the second question in which people were asked to rate how “confident” they should be in their answer given that the friend “disagrees with you”. Again, given the evidence of framing effects due to similarity in lexical content, it is plausible that this framing is likely to highlight those aspects of the case that are similarly worded. As such, they bring attention to the description of how the reader is “highly confident” in their calculation, which is described as being pretty straightforward besides the fact that the friend “disagreed about the shares”. As such, this particular formulation makes salient just how strongly the reader believes their answer to a fairly straightforward calculation is correct, and that only that friend disagrees. In this light, it is also unsurprising that people answering this question chose to maintain their high degree of confidence in their answer; after all, most of us, when we are very confident about what we believe, will not concede so easily when challenged—as testified by the endless disputes between people with different opinions.

This explanation lends support to the methodological suggestion advanced above. On this interpretation of the findings from Alexander et al. (2018), the framing effects on the Restaurant case are caused by commonalities in the wording of this scenario and the follow-up questions people responded to. This shows that, at least with respect to this case, philosophers should not ask follow-up questions which use lexical content matching those from the description of the relevant scenario, so as to avoid swaying people’s intuitions.

Lastly, it is worth mentioning how the body of empirical work on the FOR is suggestive with respect to a recently very influential proposal due to Jennifer Cole Wright (2010). She contends that attending to the perceived strength of one’s own intuitions is a reliable method for

tracking the evidentiary value of philosophical intuitions. This proposal is motivated by empirical evidence that people in experimental studies report being less confident about intuitions that are unstable—that is, those that display the worrying patterns of variation uncovered by experimental philosophers (for similar results, see Zamzow and Nichols, 2009). This suggests that philosophers should restrict their appeals to those intuitions about which people tend to report greater levels of confidence, since these are less prone to such deleterious effects.

Research into mental states with FORs shows the above methodological suggestions to be misguided. As discussed in previous sections, these studies have uncovered evidence that mental states with a high FOR—that is, responses about which people reported a high degree of confidence—are also prone to the worrying patterns of variation found in studies in experimental philosophy. Thus, people’s sense of confidence about their own intuitions is not a trustworthy method to weed out those intuitions that are subject to deleterious effects. Instead, I propose in line with the above considerations that philosophers should pay closer attention to those specific factors that have been found to sway people’s intuitions (e.g., believability of conclusions and similarities in lexical content).

In sum, I contend that we can tease out useful methodological suggestions from inquiry into the FOR that help philosophers make better use of intuitions in philosophy. Relying on recent findings in experimental philosophy, I illustrate two cases in which these suggestions prove effective at mitigating framing effects in philosophical intuitions. In light of this, I take there to be good grounds on which to deny the radical versions of the Experimentalist challenge that build on the claim that intuitions are ultimately “hopeless”. Furthermore, I propose that these considerations thus also help to articulate a moderate version of the Experimentalist challenge insofar as they inform philosophers of a set of vitiating circumstances that helps them assess their warrant for relying on intuitions in inquiry.

## **6. CONCLUSION**

I have argued that developments from research on human metacognition can help to make headway in a set of very thorny methodological disputes about the role of intuitions in philosophy. Relying on findings from this body of empirical work, I argued that a class of metacognitive states can be aptly characterized as intuitions, and that empirical findings about their workings undermine a prominent objection to the Experimentalist challenge. I then showed how inquiry into human metacognition provides us with a better understanding of the vitiating circumstances that can potentially lead intuitions astray, and demonstrated how they offer useful

methodological suggestions that help philosophers make better use of intuitions in philosophy. These arguments thus offer valuable resources that can greatly contribute to the improvement of philosophical methodology and are suggestive with respect to future inquiry into the nature and epistemology of intuitions.

## **7. ACKNOWLEDGEMENTS**

Many thanks to Lisa Bastian, Elijah Chudnoff, Eugen Fischer, Ole Koksvik, Joëlle Proust, Stefano Lo Re, and Lewis Ross for helpful discussions, and to Jessica Brown, Simon Prosser, and the anonymous referees for *Mind and Language* for their comments on earlier versions of this paper.

## REFERENCES

- Ackerman, R., & Thompson, V. A. (2017). Meta-reasoning: Monitoring and control of thinking and reasoning. *Trends in Cognitive Sciences*, 21(8), 601-617 .
- Ackerman, R., & Zalmanov, H. (2012). The persistence of the fluency-confidence association in problem solving. *Psychonomic Bulletin & Review*, 19(6), 1189–1192.
- Alexander, J., & Weinberg, J. M. (2007). Analytic epistemology and experimental philosophy. *Philosophy Compass*, 2(1), 56–80.
- Alexander, J., Betz, D., Gonnerman, C., & Waterman, J. P. (2018). Framing how we think about disagreement. *Philosophical Studies*, 175, 2539–2566.
- Arango-Muñoz, S. and Michaelian, K. (2014). Epistemic feelings, epistemic emotions: Review and introduction to the focus section. *Philosophical Inquiries*. 2(1), 97–122.
- Bengson, J. (2013). Experimental attacks on intuitions and answers. *Philosophy and Phenomenological Research*, 86(3), 495–532.
- Bengson, J. (2015). The intellectual given. *Mind*, 124(495), 707–760.
- Chudnoff, E. (2013). *Intuitions*. Oxford: Oxford University Press.
- Evans, J. S. B. T., & Curtis-Holmes, J. (2005). Rapid responding increases belief bias: Evidence for the dual-process theory of reasoning, *Thinking and Reasoning*, 11, 382–389.
- Evans, J. S. B., Clibbens, J., & Rood, B. (1996). The role of implicit and explicit negation in conditional reasoning bias. *Journal of Memory and Language*, 35(3), 392–409.
- Fischer, E. (2014). Philosophical intuitions, heuristics, and metaphors. *Synthese*, 191(3), 569–606.
- Fischer, E., & Engelhardt, P. E. (2016). Intuitions’ linguistic sources: Stereotypes, intuitions and illusions. *Mind and Language*, 31(1), 67–103.
- Fischer, E., & Engelhardt, P. E. (2017). Stereotypical inferences: Philosophical relevance and psycholinguistic toolkit. *Ratio*, 30(4), 411–442.
- Kamm, F. M. (2015). *The Trolley Problem Mysteries*. New York, NY: Oxford University Press.

- Koksvik, O. (2011). *Intuitions* (Doctoral dissertation). Retrieved from Open Research repository Australian National University. (b2638856)
- Machery, E. (2017). *Philosophy Within Its Proper Bounds*. New York, NY: Oxford University Press.
- Markovits, H., Thompson, V. A. and Brisson, J. (2015). Metacognition and abstract reasoning. *Memory and Cognition*, 43(4), 681–693. doi: [10.3758/s13421-014-0488-9](https://doi.org/10.3758/s13421-014-0488-9).X
- Nagel, J. (2012). Intuitions and experiments: A defense of the case method in epistemology. *Philosophy and Phenomenological Research*, 85(3), 495–527.
- Neys, W. D. (2006). Automatic-heuristic and executive-analytic processing during reasoning: Chronometric and dual-task considerations. *The Quarterly Journal of Experimental Psychology*, 59(6), 1070–1100.
- Nichols, S., & Knobe, J. (2007). Moral responsibility and determinism: The cognitive science of folk intuitions. *Noûs*, 41(4), 663–685.
- Petrinovich, L., & O’Neill, P. (1996). Influence of wording and framing effects on moral intuitions. *Ethology and Sociobiology*, 17(3), 145–171.
- Proust, J. (2013). *The Philosophy of Metacognition: Mental Agency and Self-Awareness*. Oxford: Oxford University Press.
- Pust, J. (2000). *Intuitions as Evidence*. New York, NY: Routledge.
- Stahl, C., Klauer, K. C., & Erdfelder, E. (2008). Matching bias in the selection task is not eliminated by explicit negations. *Thinking and Reasoning*, 14(3), 281–303.
- Stich, S. (2009). Replies. In Murph, D., & Bishop, M. (Eds.) *Stich and his critics*. Oxford: John Wiley & Sons.
- Swain, S., Alexander, J., & Weinberg, J. (2008). The instability of philosophical intuitions: Running hot and cold on truetemp. *Philosophy and Phenomenological Research*, 76(1), 138–155.
- Thompson, V. A., & Johnson, S. C. (2014). Conflict, metacognition, and analytic thinking, *Thinking and Reasoning*, 20(2), 215–244.
- Thompson, V. A., Evans, J. S. B., & Campbell, J. I. (2013a). Matching bias on the selection task: It’s fast and feels good. *Thinking and Reasoning*, 19(3–4), 431–452.
- Thompson, V. A., Turner, J. A. P., & Pennycook, G. (2011). Intuition, reason and metacognition.

*Cognitive Psychology*, 63, 107–140.

Thompson, V. A., Turner, J. A. P., & Pennycook, G. (2013b). The role of answer fluency and perceptual fluency as meta-cognitive cues for initiating analytic thinking. *Cognition*, 128, 237–251.

Thomson, J. J. (1985). The trolley problem. *The Yale Law Journal*, 94(6), 1395–1415.

Tsujii, T., & Watanabe, S. (2010). Neural correlates of belief-bias reasoning under time-pressure: A near-infrared spectroscopy study. *NeuroImage*, 50, 1320–1326.

Weinberg, J. (2015). Humans as instruments: Or, the inevitability of experimental philosophy. In Fischer, E., & Collins, J. (Eds.) *Experimental philosophy, rationalism, and naturalism. rethinking philosophical method*. New York, NY: Routledge.

Weinberg, J. M. (2007). How to challenge intuitions empirically without risking skepticism. *Midwest Studies in Philosophy*, 31(1), 318–343.

Wright, J. (2010). On intuitional stability: The clear, the strong, and the paradigmatic. *Cognition*, 115(3), 491–503.

Zamzow, J. L., & Nichols, S. (2009). Variations in ethical intuitions. *Philosophical Issues*, 19(1), 368–388.