William Ramsey
Department of Philosophy
University of Nevada, Las Vegas

THE COGNITIVE SOURCE OF PHILOSOPHICAL INTUITIONS: TROUBLES FOR THE ETIOLOGICAL PROJECT*

Abstract: *The advent of experimental philosophy has generated a renewed interest in the nature of philosophical intuitions. This has led many to assume that understanding the psychological processes that generate philosophical intuitions will provide a much needed source of answers for various questions about their nature. It is widely assumed that if we are to gain a good sense of the reliability of philosophical intuitions (and the degree to which experimental philosophy challenges their role in philosophy) it is critical to learn more about the sort of cognitive faculties and processes that produce them. However, here, I'm going to present a more pessimistic outlook about what we should expect from such a project in the case of philosophical intuitions. More specifically, I'll suggest that there are two central problems that prevent a deeper understanding of the cognitive source of intuitions from doing much of what we would like it to do. The first problem is that for an important range of philosophical intuitions, it is doubtful that there is a stable set of cognitive processes and mechanisms generating them that is sufficiently universal to allow for the sort of generalizations we would like. Philosophical intuitions are likely created by an assortment of varying psychological operations, and that learning about how a particular intuition is created is unlikely to tell us much about how other intuitions are formed. In short, there is no faculty of intuition in any meaningful sense. I call this the "Etiological-Diversity Problem". The second problem stems from the fact that one of the primary things we want knowledge of the intuition-producing cognitive machinery to tell us about is the reliability of the intuitions they produce. However, this expectation gets the normal order backwards. The evaluation of cognitive processes or mechanisms typically requires a prior evaluation of the normative status of the states it produces in different contexts, distinguishing proper functioning and competence from improper functioning and performance errors by virtue of the quality of the mechanism's output. It is far less clear how this order can be reversed. Consequently, it is hard to see how an understanding of intuition-producing mechanisms will allow us to evaluate the epistemic status of philosophical intuitions without already having such an evaluation in hand. I call this the "The Secondary Calibration Problem". The upshot is that even though I believe there is much to learn about the etiology of intuitions that will be useful, these problems severely hamper our ability to answer many of the central questions we have about the nature and trustworthiness of philosophical intuitions.*

Keywords: *etiological project, cognitive program, etiological diversity, calibration problem*

I. Introduction

The discoveries and theoretical developments associated with experimental philosophy have turned new attention to philosophical intuitions – upon their nature, their reliability, their stability, and a host of other dimensions. With this recent interest, it has become increasingly common for authors to suggest that understanding the psychological processes that generate philosophical intuitions will provide a much needed source of answers for various questions about their nature. More specifically, it is generally assumed that if we are to gain a good sense of the reliability of philosophical intuitions (and the degree to which experimental philosophy challenges their role in philosophy) it is critical to get a firmer grasp of the sort of cognitive faculties that produce them. For example, Helen De Cruz states that ". . . understanding the etiology of philosophical intuitions is vital for making headway in debates on their evidential value" (De Cruz, 2015, p. 235). Similarly, Brian Talbot tells us that the best way to address concerns about philosophical intuitions is to "develop a general and systematic understanding of how intuitions work: where they come from, how they are generated, what they are and are not based on, what factors affect them" (Talbot, 2009, p. 159). Echoing these themes, Joshua Alexander asserts that, "[w]hat is needed, then, is a better understanding of the cognitive processes involved in the production of our philosophical intuitions. By coming to better understand what intuitions are, where they come from, and what factors influence them, we can better understand what role they can play in philosophical practice" (Alexander, 2012, p. 3).[1]

For the most part, I am sympathetic with the idea that a good way to enrich our understanding of some psychological state or condition is to attend to its etiology. If we want to know the nature of some mental state and its relation to the world, then it certainly makes sense to attend to how such states are formed – to the cognitive machinery and conditions that generate them. However, despite the prima facie reasonableness of many such projects in psychology, I'm going to present a more pessimistic outlook about what we should expect from such a project in the case of philosophical intuitions. More specifically, I'll suggest that there are two central problems that prevent a deeper understanding of the cognitive source of intuitions from doing much of what we would like it to do. The first problem is that for an important range of philosophical intuitions, it is doubtful that there is a stable set of cognitive processes and mechanisms generating them that is sufficiently universal to allow for the sort of generalizations we would like. That is, for an important class of philosophical intuitions, they are likely created by an assortment of varying psychological operations, and that learning about how a particular intuition is created is unlikely to tell us much about how other intuitions are formed. In short, there is no faculty of intuition in any meaningful sense. I

---

1   For further examples of this perspective, see Allman and Woodward, 2008; Fischer and Collins, 2015; Knobe and Nichols, 2008; Knobe and Prinze, 2008; Scholl, 2007.

call this the "Etiological-Diversity Problem". The second problem stems from the fact that one of the primary things we want knowledge of the intuition-producing cognitive machinery to tell us about is the reliability of the intuitions they produce. However, this expectation gets the normal order backwards. The evaluation of cognitive processes or mechanisms typically requires a prior evaluation of the normative status of the states they produce in different contexts. We typically distinguish proper functioning and competence from improper functioning and performance errors by virtue of the quality of the mechanism's *output*. It is far less clear how this order can be reversed – how we can use understanding of the functioning of a cognitive process to then evaluate the nature of the cognitive state it generates. Consequently, it is hard to see how an understanding of intuition-producing mechanisms will allow us to evaluate the epistemic status of philosophical intuitions without already having such an evaluation in hand. I call this the "The Secondary Calibration Problem". The upshot is that even though I believe there is much to learn about the etiology of intuitions that will be useful, these problems severely hamper our ability to answer many of the central questions we have about the nature and trustworthiness of philosophical intuitions.

To show all this, the next section will fill in some of the background on what I will call the "etiological project"[2], briefly looking at some of the programs of cognitive research where it has been pursued in understanding the way philosophical intuitions are generated. Then, in Section III, I will present the etiological-diversity problem and explain how it is likely to undermine our ability to make significant generalizations about the way philosophical intuitions are produced. In Section IV, I will present the secondary calibration problem and explain how it poses a barrier to our epistemic evaluation of intuitions. Section V offers a brief conclusion and some recommendations about how to move forward in light of these considerations.

II. Investigating the Etiology of Intuitions: The Why and the How

The term "philosophical intuition" is widely recognized as having an unclear extension and refers to a range of different types of states and attitudes (Nado, 2013). To focus our discussion, I will restrict my analysis to the sort of intuitions generated by the consideration of a philosophical vignette or thought experiment, the scenarios that comprise what is often referred to as "the method of cases". Famous examples commonly discussed in the literature include the intuitive judgments prompted by Gettier's 10 Coin Case, Kripke's Gödel cases, Thompson's violinist case, and so on. We can refer to this class of intuitions as Thought Experiment Intuitive Judgments[3]

---

2   This has also been referred to as "the sources project" (Pust, 2017) and the "cognitive program" (Sytsma and Livengood, 2016).

3   The term 'judgment' is ambiguous, in that it refers to both a process (the judging) and the end result of the process (the resulting judgment). Here I am using the term in the

or "TEIJ"s for short. Even with this restriction, there are still debates about the psychological nature of such states. Some regard them as a type of belief (e.g., Lewis, 1983), while others suggest they are an inclination or disposition to believe (Sosa, 1998; Earlenbaugh and Molyneux, 2009), while still others argue that they are a distinctive type of "intellectual seeming" or a distinct type of propositional attitude (Bealer, 1998). The latter strikes me as the most plausible perspective, although not a lot of what I will say will depend upon this. Writers have also emphasized the distinction between the psychological state – the intuiting – and the content of such a state – the intuited (Lycan, 1988). Our discussion here will focus upon the psychological processes and mechanisms that produce the former.

Let's begin by stepping back and considering why we want to study the etiology of intuitions in the first place, and then we can consider how such a study is likely to go. As suggested by the quotations in the introduction, the desire to learn about the cognitive processes and mechanisms that create philosophical intuitions is grounded in the assumption that doing so will help us answer a number of important questions about their nature. A couple of central questions we assume studying the etiology of intuitions will help us answer are these:

1. Are the cognitive processes and mechanisms that produce intuitions trustworthy – do they produce accurate representations of facts concerning philosophical matters, and if so, how?

2. Insofar as such mechanisms occasionally go awry, what sorts of conditions and situation are likely to lead to their improper operation and thus to the production of faulty intuitions, and how can these conditions and situations be avoided?

The linking of these sorts of questions and empirical work on the etiology of intuitions is, at least initially, easy to understand. If you want to know about the reliability of a cognitive mechanism, then you need to understand its actual nature – its basic operation, the sub-systems involved, the sort of computational processes involved, its responsiveness to mind-independent reality, and so on. If you want to know how it properly functions, and to be able to distinguish proper function from malfunction, then, at the very least, you need to learn about its normal operations. Moreover, learning about the functioning of a faculty is often a critical step in learning about the nature of the sort of cognitive state it generates.

There are couple of important things to notice about these questions. The first is their clearly normative nature. Notions like trustworthiness or proper and improper operation demand more than a straightforward mechanical description of the system's operations; they demand that we make value assessments that rank some types of functioning as superior to

---

latter sense – referring to the state produced, the relevant intellectual seeming that such and such is the case.

others. We want to gain knowledge about the processes and mechanisms that create TEIJs because we think such knowledge can tell us something about the *quality* of those intuitive states. As we will see below, establishing this normative dimension for the cognitive machinery responsible for an important class of intuitions is going to be deeply problematic. The second thing to note is that questions like these are related; answers to some clearly depend upon the answers to others. Questions about malfunction or about flawed performance presuppose that we can ascertain when a faculty is functioning properly, and that clearly requires knowledge about the process whereby it generates appropriate outputs. To some degree, questions like these stand or fall together. Difficulties with answering any one of them will likely raise significant problems for answering others.

So much for the why – what about the how? How should the etiology of intuitions be investigated? While some have suggested that intuitions can be adequately investigated using introspection or some sort of a priori reasoning, there is a growing consensus that the production of philosophical intuitions is an empirical matter and should be investigated by the relevant areas of cognitive science (Talbot, 2009). Like any other sort of cognitive state, philosophical intuitive judgments are first and foremost types of psychological phenomena that should be treated as such. Still, this constraint allows for a very broad range of possible research programs and approaches. Pertinent research has come from (and likely will continue to come from) an array of different research programs, including everything from low-level neurological studies and imaging to much more abstract theorizing about computational processes. Indeed, just a casual survey reveals several different research approaches that have yielded accounts of the cognitive processes responsible for intuitions. While these approaches are not completely distinct and often overlap a great deal, they nevertheless involve their own presuppositions, conceptual apparatus and over-arching perspectives. Three prominent research approaches to understanding the cognitive source of TEIJs are: 1) the concept application/categorization approach, 2) the dual-process framework and, 3) the heuristics approach. It will help to briefly consider each of these.

A. The Concept Application/Categorization Framework

It is often assumed that for TEIJs, the judgment stems from the application of a philosophical concept to a particular situation. The content of nearly all TEIJs can be characterized as having the form: 'the X in this case is an instance of Y' (e.g., the belief in this case is an instance of non-knowledge; the act in the case in question is an instance of immorality; the behavior in this case is an instance of free action, and so on). As Goldman puts it, intuitions "are what might be called *classification* or *application* intuitions, because they are intuitions about how cases are to be classified, or whether various categories or concepts apply to selected cases" (2007, p. 4). Thus, one clear

area of research used to understand the production of TEIJs is psychological research on concept application. Within this thriving area of research, there have been three central theoretical paradigms about concepts and how they are employed in categorization judgments: the prototype paradigm, the exemplar paradigm and the theory-based account. Each of these theoretical frameworks has demonstrated success in explaining important findings with regard to the way we categorize entities, kinds, properties, events, and the various other things concepts are about. For example, within the prototype paradigm, categorization is usually driven by matching features possessed by the target with weighted features that comprise the concept. This approach has had considerable success in explaining what has come to be referred to as "typicality effects" (Rosch, 1978), such as our tendency to rank members of a kind as more typical than others, (Smith and Medin, 1981). By contrast, with the exemplar tradition concepts are stored familiar instances encountered (Barsalou, L. W., Huttenlocher, J., & Lamberts, K., 1998), so that, say, one's concept of DOG would be represented by a familiar dog stored in memory, like the family pet Fido. This helps explain the speed with which we categorize atypical but familiar instances. The third tradition – the theory-based account – claims that concepts are represented by tacit theories about the concept's target (Keil, 1989). This model explains the various ways in which our categorization activity looks more like the application of a theory rather than a comparison to some internally stored representation.

On the concept application approach, when imagining the scenario presented by a philosophical thought experiment, we classify something or someone in the scenario as qualifying (or failing to qualify) as an instance of a philosophical property by making a categorization judgment using something like a prototype, exemplar or tacit theory. We will look more closely at how these three models of concepts could be employed to explain philosophical intuitions in the next section. As a preview, there is good reason for thinking that all three of these models of concept application are employed by our brains. Thus, even if we restrict the etiological project to the concept application/categorization approach, it is likely that the way in which we form TEIJs varies dramatically between different mechanisms and operations.

B. The Dual Process Approach

Another area of research associated with the production of intuitions is the "Dual Process" or "Dual System" account (Evans, 2010; Evans and Frankish, 2009; Evans and Over, 1996). According to this view, there are two types of processing architectures in the brain. On the one hand, there are judgments that involve explicit conscious reasoning and typically result in the formation of a belief. This is what happens when, say, someone deliberates on the premises of an argument and consciously forms a conclusion. This introspectively accessible process is said to be performed by "System 2" in the brain. This is contrasted with "System 1" psychological processes that are

unconscious, more automatic, and typically outside of our control. Several authors have noted that it is likely that our intuitive responses to hypothetical cases in philosophy are generated by this latter, System 1 type of processing (Weinberg and Alexander, 2014; De Cruz, 2015). The processes that generate TEIJs are not introspectively accessible and appear to be quite automatic – we hear the scenario and just find that it strikes us a certain way. This has led at least some to propose investigating the psychological source of philosophical intuitions by using the dual-processing approach.

For example, DeCruz (2015) has employed psychologist Robert McCauley's account (2011) of two different sorts of Type 1 processing to explain a variety of philosophical intuitions, particular those involving teleology and epistemology. As De Cruz notes, McCauley distinguishes between what he refers to as "maturationally natural cognition" and "maturationally practiced cognition". The former involves Type 1 processing that spontaneously results from natural development, and requires no special training, such as learning to speak one's native language or ascribing mental states to others. The latter involves Type 1 processing that involves a distinct skill and requires some degree of training or practice, such as riding a bicycle or playing the guitar. De Cruz argues that both sorts of cognition likely give rise to philosophical intuitions. For instance, she suggests that intuitions involving teleology and knowledge possession arise from maturationally natural cognition (because these come about spontaneously without any special training) whereas intuitions about, say, how Hume would respond to a certain line of inquiry would be produced by maturationally practiced cognition (because it requires immersion into Hume scholarship). This is just one of the ways in which dual processing theories can be used to account for the production of philosophical intuitions.

C. The Heuristics Approach

A third area of research that has made a significant contribution to the etiological project is the important psychological work on various everyday heuristics that we use in a wide range of endeavors (Kahneman et al., 1982; Gigerenzer et al., 1999). Heuristics are often characterized as processing short-cuts or rules of thumb that are employed unconsciously to help us solve various problems. They tend to be faster and easier to use than more complicated calculations, but they can also be unreliable and can lead to errors. A classic example is the "recognition heuristic". Here, subjects employ the recognizability of some instance as an indicator of some other, less accessible property it might possess. For example, when asked which city is larger, it has been shown that people tend to use their recognition of the city's name as a guide to the size of the city – the more recognizable, the larger the city is assumed to be. This seems like a reasonable strategy since there is a rough correlation between the fame of a city and its size; however, it can lead

subjects to make faulty inferences, even when they are provided with further information in particular cases suggesting that the recognizability of the city should be discounted (Gigerenzer et al., 1999).

One area of philosophical investigation that has explored the role of heuristics in producing philosophical intuitions is moral philosophy and psychology. Several authors have suggested that our moral intuitions are produced by moral heuristics that are employed in our moral judgments and often allow us to assign moral responsibility or determine a course of action. For example, Sinnott-Armstong et al. (2010) develop an account of moral heuristics and intuitions based upon Kahneman and Frederick's (2005) notion of unconscious attribution substitution. With unconscious attribution substitution, people unconsciously determine whether or not something has a certain attribute, F, by ascertaining whether or not it has some other attribute, G. People are typically unaware of substituting the heuristic attribute for the target attribute, and the substitution clearly involves some sort of tacit assumption about a correlation between the two. It is a useful heuristic because the target attribute is often difficult to detect, while the heuristic attribute is easier to reveal. Sinnott-Armstrong et al. go on to develop a sketch of how such a heuristic could give rise to moral intuitions. Because the moral rightness or wrongness of an act can often be very difficult to discern, they suggest that people substitute some other attribute, like what the majority is doing, or whether or not the act produces unpleasant feelings, to ascertain moral propriety. On this account, when presented with a thought experiment in ethics, the audience would intuitively judge what ought to be done by considering something else, like how the scenario makes one feel.

Summary

This is not intended to serve as an exhaustive overview of the only research programs that hold considerable promise of telling us about the source of TEIJs. To the contrary, it is intended to help illustrate just how much theoretical diversity exists with regard to the etiological project. And while these approaches are not always mutually exclusive, it is still true that, given the undeniable differences in these research agendas, this variance in theorizing is likely to contribute to the problem I will raise in the next section.

III. The Etiological-Diversity Problem

Before explaining the etiological-diversity problem, it will help to consider a couple of ways in which diversity regarding intuitions is _not_ the worry I'm concerned about. One sort of diversity was mentioned in Section II – that a very broad range of different sorts of psychological states get referred to as 'intuition', including the seeming truthfulness of logical claims, semi-informed hunches, the puzzling nature of a paradox, and so on. But this diversity can be

made far less problematic by simply stipulating that we restrict our analysis to one specific type of intuition, as I proposed in Section II. Thus, we can limit our focus to the sort of intuition that has played such an important role in theory development in philosophy – the immediate, spontaneous reaction to a philosophical thought experiment, what I've been calling TEIJs. Of course, TEIJs may still involve a range of different types of mental states, in which case my pessimism would only be reinforced. But for now, for the sake of this discussion, I'm going to assume a fair degree of uniformity among TEIJs, qua-mental state type. The second type of non-problematic diversity pertains to the fact that, as with nearly all faculties, the TEIJ-producing process probably involves a "boxology" of different sub-systems and modules that collectively generate the intuitive reaction. The mere existence of an array of contributing sub-modules poses no problem to developing a robust theory of how a given faculty works – indeed, spelling out the organizational mechanics or "flow-chart" of different cognitive process often provides one of the most satisfying explanations of how a cognitive task is performed.

If these forms of diversity do not pose a problem for the etiological project, then what sort of diversity does? Consider two different types of psychological states. For one type, there is something close to self-contained cognitive system or module that is responsible for creating the state in question. For example, our perceptual systems more or less work in a predictable and systematic manner to produce specific perceptual states such as visual or auditory experiences. We can study the auditory system and learn about its basic functioning – about the relevant parts of the brain that are involved and the sort of computational operations employed – and thereby gain a great deal of understanding about how hearing works. This understanding yields helpful predictions and generalizations because, although it is incredibly complicated, it is nevertheless the same core auditory system that is working throughout the wide array of conditions in which we use our hearing to navigate the world. While there are a variety of different stages and sub-systems involved, most of the time hearing involves the same sub-systems. Hence, there is a faculty of hearing that lends itself to comprehension, and thus allows for useful generalizations and predictions about the etiology of auditory experience.

However, not all psychological states have such an unvaried, self-contained faculty that is responsible for their production. Consider beliefs. Given the remarkably diverse array of cognitive processes and mechanisms that can generate beliefs – perception, memory, inference, introspection, etc. – it would clearly be a fundamental mistake to launch an inquiry into the "faculty of belief". That is, given the extreme diversity of cognitive processes that produce beliefs, we should not expect to find an account of belief production that will yield the sort of predictions and generalizations that we find with perceptual states. Learning about how beliefs are formed via a process of inference is not going to tell us very much about the way beliefs are

formed via perceptual experience or recall. Of course, this doesn't mean there is no value to studying the cognitive etiology of beliefs, or that we can't gain useful information about these diverse processes of belief formation. But in the case of beliefs, it is clearly a mistake to ask questions such as whether or not *the* process that generates beliefs is reliable or improves with training or tends to malfunction in certain precise conditions. It is a mistake to ask such questions because there is no unique, uniform process that generates beliefs.

So, then, what about TEIJs? Traditionally, many have assumed that the proper model for understanding the etiology of intuitions is something like the perceptual model just presented. Many philosophers have treated rational intuition as if it was the a priori analogue of our senses, providing basic evidence about hidden philosophical truths, albeit in a non-sensory way (see, for example, Sosa, 1998; Hales, 2012). The intuition faculty is thereby treated as doing the same sort of work in philosophy that observation does in science – as providing a kind of access to extra-mental facts about ethics, epistemology, mind, language, and so on.

Naturalistically inclined philosophers, including those in experimental philosophy, have come to reject the idea that intuitions are produced by cognitive systems that provide a window or insight into the realm of abstract philosophical truths. They are instead committed to explaining the production of intuitions along the lines discussed in the last section – as a process of concept application or as involving the use of various heuristics. Yet, while there has been a rejection of the perception-like model, it is less clear that naturalistic philosophers have abandoned the related idea that there is a stable faculty or module that generates intuitions about particular philosophical matters. For example, Shieber (2012) suggests that there are domain-specific, hard-wired, functionally distinct modules that are similar to a linguistic processing module and that generate intuitions about the different domains of philosophy, such as ethics, epistemology and metaphysics. While Shieber's "Modularity Model of Intuition" stipulates that there are different modules corresponding to the different areas of philosophy, it implies that the cognitive structures delivering intuitions in each area are relatively stable, fixed and uniform.[4]

But it is far from clear that an even limited sort of uniformity thesis is correct. Instead, as I'll now argue, when we look closely at the way the etiological project is actually carried out, there is good reason to suppose that the array of operations and sub-mechanisms producing philosophical intuitive reactions, even in specific areas of philosophy, vary and diverge significantly. That is, the expectation that we will discover a stable module or set of modules that systematically generate TEIJs is overly optimistic. Given the way the research is unfolding, we should instead anticipate a mixed bag of varying procedures and mechanisms that produce TEIJs and that will preclude substantial generalizations and predictions.

---

4   Similar perspectives have been suggested about epistemological intuitions by Nagel (2012) and moral intuitions by Hauser, Young and Cushman (2008).

As we saw in Section II, the range of different research programs that have contributed theories about the cognitive etiology of TEIJs is fairly broad. These different research programs, with their own theoretical posits, principles and approaches, create considerable theoretical diversity despite overlapping to some degree. McCauley's Type 1 account of cognition as a source of intuition involving 'maturationally natural' and 'maturationally practiced' cognitive processes is very different from, say, a prototype account of how the judgment is produced that appeals to a typicality ranking, or an account that invokes a specific heuristic. Of course, *theoretical* diversity does not entail *processing* diversity or *mechanism* diversity. Still, this diversity in the theoretical terrain at least suggests that we should not expect a single coherent etiological account any time soon. Moreover, there is reason to believe that these different theoretical frameworks are enjoying success because they are actually describing different mechanisms and processes that produce TEIJs. In other words, insofar as they disagree, these competing theoretical frameworks may nevertheless be correct because they are capturing multiple types of operations and processes that are actually devoted to generating different intuitions in different circumstances.

Suppose we limit our evaluation to just one of these research approaches, and focus solely on the concept application/categorization framework. Recall that with this approach, TEIJs are identified with categorization judgments, produced by the application of philosophical concepts. The processes that generate philosophical intuitions are the psychological processes associated with the employment and application of a concept. But even if we focus just upon the concept application perspective, there is good reason to expect a considerable amount of etiological diversity. As we noted in Section II, there are three popular models – prototype, exemplar, and theory models – about what concepts are, how they are stored and retrieved, and how they allow us to classify things. Of course, within each of these paradigms there are theoretical variations and sub-variants endorsed by different researchers. For instance, there are different versions of prototype accounts that involve diverse ways in which item features are invoked for categorization (Murphy, 2002). This is perfectly normal for thriving research programs where an array of alternative theories compete for consensus through predictive and explanatory success. Yet even if we ignore the sub-variants within in each paradigm, competition among the three accounts of concept application has not yielded a clear winner. Instead, the different theories seem to be enjoying considerable success because of different robust findings that uniquely support each particular model. For instance, the easy learnability and classification of items that possess more prominent and common features supports most feature-based prototype accounts. However, the finding that we sometimes classify more quickly atypical members that are nevertheless similar to those we have been exposed to through personal experience undermines the prototype view, but instead supports the exemplar theory.

Various other findings that suggest classification does not involve any sort of the comparison process lend credence to the theoretical account. And, of course, there are findings that serve as evidence for theories of concepts that are different yet from these three (Murphy, 2002, Machery, 2009).

This diversity of successful theorizing has led some, in particular Edouard Machery and Daniel Weiskopf, to endorse what is sometimes known as the "heterogeneity hypothesis" or "pluralism" about concepts (Machery, 2009; Weiskopf, 2009). The heterogeneity/plurality hypothesis is simply the view there is no single correct answer to what it is for something to be a concept. For most items (kinds, properties, relations, events, etc.) we actually possess different kinds of concepts, particularly those represented in the manner suggested by the three models presented above. In other words, on the heterogeneity/plurality hypothesis, there is psychological variance not only in terms of what the concept is about (e.g., dog vs. cat), but also in terms of the sort of knowledge structure that is involved. When I go to classify something as an instance of, say, a dog, in certain contexts I'll do so by employing a prototype representation of a dog, while in other contexts I'll do so by invoking an exemplar (stored memory of familiar dog) or some deeper theoretical knowledge about dogs. There is no one account that is right and that can yield straightforward generalizations about the process of categorization. In other words, on the heterogeneity/plurality hypothesis, the cognitive nature of concept application/categorization is varietal and (duh!) heterogeneous.

A robust defense of the heterogeneity/plurality thesis about concepts would take us too far afield, so I will invite readers to examine the arguments presented by Machery and Weiskopf and decide for themselves. Most of these arguments strike me as absolutely convincing and present the most compelling way of making sense of the divergent psychological findings regarding the nature of concepts.[5] And while Machery goes so far as to recommend abandoning talk of concepts altogether (since the heterogeneity hypothesis implies there is no conceptual natural kind), we do not need to embrace concept eliminativism for the heterogeneity/plurality hypothesis to have profound implications for the etiological project. If it is correct, the heterogeneity/plurality hypothesis would mean that insofar as philosophical intuitions prompted by thought experiments are psychologically generated by the application of philosophical concepts, their production would involve, as a group, a very diverse range of different cognitive processes.[6] It wouldn't

---

5   I will only mention that at least some alternative accounts of the divergent findings, such as various types of hybrid accounts (for example, Osherson and Smith, 1981) partially agree with the heterogeneity/plurality thesis that different bodies of knowledge and processes can be involved in the application of a concept. For example, in Osherson and Smith's account, we might classify something by using a definition in some contexts and by using a prototype in other contexts. With such an account, there would also be diverse processes that generate TEIJs.

6   Thus, my claim in this section is a conditional one: if the heterogeneity/plurality thesis is correct, it would mean that even if TEIJs are solely produced by processes of concept application, this would still involve a diverse range of very different processes generating TEIJs.

simply mean that the cognitive process that generates an intuitive reaction to a Gettier-type case might involve one type of concept, say a prototype concept of knowledge, while the process that results in an intuition about a Kripke-style Jonah case would involve a different type of concept, say a tacit theory about reference. It would also mean that the conceptual machinery employed in response to one sort of Gettier case could be substantially different than the conceptual machinery employed in a different sort of Gettier case. Indeed, since several different factors can trigger the employment of one type of concept rather than another, it could entail the retrieval of different types of concepts in different people with regard to the same Gettier case, or different types of concepts in the same person in response to the same case at different times in different settings. Learning that subjects A, B and C employed a certain sort of prototype in categorizing Gettier's 10 Coins Case as non-knowledge would be perfectly compatible with subjects D, E and F employing an exemplar concept in response to that case, or A, B and C employing an exemplar concept in response to a Fake Barn case. The amount of intuitional etiological diversity allowed by the heterogeneity/plurality hypothesis is such that it would undermine the hope of making broad generalizations about how, psychologically, humans respond to philosophical thought experiments. Even if we could somehow determine that, say, a particular process involving a prototype concept of knowledge yielded a reliable intuition about a Gettier case, we would not be able to generalize that assessment to Gettier intuitions more broadly given that it is unlikely that such a process is invoked universally. And remember, this is the range of etiological diversity that exists even after stipulating that our intuitive responses to philosophical cases are entirely due to the application of concepts. Considering that it is more likely that TEIJs are sometimes produced by processes other than concept application, such as the employment of a heuristic or some other form of Type 1 mechanism, then my pessimism about the etiological project providing useful generalizations looks even more warranted.

So far I've addressed the prospects of the etiological project from the perspective of some of the different research programs devoted to explaining the generation of philosophical intuitions. My claim has been that from the perspective of these different research agendas, it is highly doubtful that we will eventually end up with an all-encompassing, general account of intuition etiological processes and mechanisms or modules that can provide very helpful generalizations or that can answer the sort of broad questions many have about TEIJs. We are instead likely to find a diverse array of idiosyncratic and highly specific etiological processes that apply only to particular intuitive judgments in particular situations. This verdict is substantially reinforced when, instead of looking at research paradigms broadly construed, we look instead at the way investigators try to account for specific intuitive judgments that are reactions to specific thought experiments. When we look at proposals about how individual sorts of TEIJs are produced, we find very particular, idiosyncratic accounts positing operations that make sense only

as theories about the specific sort of intuitive judgment in question. Theories about the way, say, certain intuitions about moral propriety are produced posit such detailed and specific machinery that there is very little chance that the same sort of machinery has much to do with the way intuitions about epistemology or metaphysics are produced. Indeed, even within particular areas of philosophical specialization, like ethics or epistemology, the cognitive processes invoked to explain specific philosophical intuitions prompted by one sort of case often have little in common with the cognitive processes invoked to explain other philosophical intuitions about other sorts of cases in the same area. In a sense, this shouldn't be terribly surprising. Researchers construct models that are fine-tuned to explain idiosyncrasies and specific noteworthy findings associated with individual TEIJs. In the process, they typically propose functional elements and operations that would generate the particular observed effects associated with the intuitive reaction, and that make sense of only certain distinctive intuitional phenomena.

To see this better, consider one of the most active areas of intuition etiological research – work regarding the attribution of intentional action and on what has come to be called the "side-effect effect". Joshua Knobe (Knobe, 2004, 2008) has shown that people's intuitions about whether or not an act was performed intentionally is often strongly influenced by their moral evaluation of the act. Subjects are presented with two scenarios: one where the chairman of company adopts a policy for economic reasons but that he knows will also harm the environment, and one, (worded in nearly identical language) where he knows it will have the side-effect of helping the environment. Remarkably, 82% of the subjects intuited that the vice president intentionally harmed the environment, but only 23% said he intentionally helped the environment. It seems that the subjects' moral assessment of an act (harming as opposed to helping) strongly influenced their views on the actor's intentions.

Because this is a surprising finding about an intuitive response, there has been considerable speculation about the underlying psychological process that creates this TEIJ. For our purposes, what matters is the specific nature of cognitive models that have been proposed. For example, Knobe has argued that the reaction in the harm case is due to a subtle interplay of perceived features of the case that (unconsciously) play a critical role in the formation of the intuition of intentionality. The process is one where the subject first determines if the behavior is bad or good, and then assigns intentionality depending on the presence of differing features, such as the result being foreseeable by the agent or involving some degree of effort, depending on the specifics of the case. On this etiological model, "people's intentional action intuitions tend to track the psychological features that are most relevant to praise or blame. But – and this where the moral considerations come in – different psychological features will be relevant depending on whether the behavior itself is good or bad" (Knobe, 2008, p. 144).

The critical thing to note here is how specific this account is to the particular sort of TEIJ that is being explained. This account makes sense only with regard to intuitions about whether or not an act is performed intentionally by someone in a specific vignette. The account has little value in explaining the production of other philosophical intuitions, such as intuitions about knowledge, personal identity, reference, or even intuitions about deliberate action that are prompted by scenarios significantly different from the harm/help scenarios. That's because the model is designed to account for a very particular intuition, and is in no way intended as some sort of general purpose model. Indeed, even within Knobe's account, there is considerable variance. As Knobe puts it, "…a given feature may be highly relevant to the praise or blame the agent receives for one behavior while remaining almost entirely irrelevant to the praise or blame the agent receives for another, somewhat different behavior…People's intentional action intuitions seem to exhibit certain flexibility, such that they look for different features when confronted with different behaviors…" (2008, 143–144). In other words, the specific process that generates an intentionality intuition is going to vary from case to case.

When we look at the models designed to explain other philosophical intuitions, we find the same sort of idiosyncratic functionality (see for example, Arico, et. al., 2011; Sytsma and Machery, 2009). Between these accounts, we see very little overlap, except in the boring sense in which they all attempt to explain an intuitive response by appealing to mechanisms and processes that "make sense" of that particular intuition. Beyond that, these accounts have about as much in common as psychological accounts of inference and short-term memory – both processes that generate beliefs. For most experimental philosophers engaging in the cognitive program, explanatory breadth and broad-ranging principles of the sort that would allow noteworthy generalizations are neither essential nor expected. Given this trend in explaining where TEIJs come from, it seems quite unlikely that we will get any sort of unified theory about a faculty that is *the* source of philosophical intuitions.

To wrap up this section, consider again one of the central sort of questions it is hoped the etiological project will answer: "Are the cognitive processes and mechanisms that produce philosophical intuitions in response to a thought experiment adequately reliable?" It is becoming increasingly clear that the answer no doubt depends upon *which* philosophical intuitions and *which* cognitive process and mechanisms you are talking about. Indeed, the answer may depend upon which cognitive processes and mechanisms producing which intuitions, in whom, in what sort of setting. There really is no reason to think that we will eventually develop a uniform account of something like an intuition faculty that can answer broad questions about the general nature of TEIJs.

IV. The Secondary Calibration Problem

In the last section, I raised doubts about our ability to find general answers to important questions about the etiology of our intuitive reactions to philosophical thought experiments – doubts based upon signs pointing to the extreme diversity of processes and mechanisms that can give rise to TEIJs. In this section, I'm going to raise another problem that would exist even without this high degree of etiological diversity. As we saw in Section II, the questions we seek to answer about intuitions by investigating their etiology have a normative dimension to them. Are the intuitions produced trustworthy? In what sort of contexts are they likely to go awry in the production of intuitions? The etiological project is at least partially based upon the hope that a deeper understanding of the cognitive processes and mechanisms that generate intuitions will help us to answer these questions.

This hope is largely motivated by the recognition of a difficulty in evaluating the truthfulness of philosophical intuitions. As a number of authors have pointed out, perhaps most notably Robert Cummins (1998), we do not have an independent method for assessing the truthfulness of the contents of these intuitive reactions; we only have the intuitive reactions themselves. It may seem true that in Gettier's 10 coins case Smith's belief is not knowledge, but, ironically, it is far from clear how we can know if this intuitive reaction is correct. As Cummins notes, processes and mechanisms that deliver information need to be properly calibrated, and proper calibration requires independent access to the sort of facts or state of affairs that the information is about. But in most cases, there is no independent access to the philosophical matters that our intuitive reactions are about. This is somewhat unique to TEIJs. If something perceptually seems a certain way to us, like the lines appearing to be of different length in the Mueller-Lyer illusion, there are strategies we can employ to establish that this seeming is mistaken, that the lines are actually the same length (we can, for example, measure them). If we want to test the reliability of memories in certain contexts, we can do things (like, say, use audio-video recordings) to provide an independent mode of access to the relevant source of information, and then compare this to the memories. But what do we examine to test the sort of metaphysical, epistemological and ethical intuitions that are generated by the presentation of a thought-experiment? What can we do to test whether the common reaction to a Gettier case reveals a deep fact about knowledge, as opposed to a widespread delusion about knowledge? With regard to the way things intuitively seem to us in response to many philosophical thought experiments, there is no real strategy for assessing those intellectual seemings beyond the philosophical theories that the intuitions are employed to support or challenge. Reflective equilibrium is often presented as a process whereby we *can* do something like calibration – weighing our commitment to the intuition against theoretical commitments with which they conflict.

Yet, as Cummins and others have noted, we very rarely reject an intuition because it conflicts with a philosophical theory. Instead, we almost always use the intuition as compelling evidence that a given theoretical stance is false. This has come to be known as the "calibration problem" with regard to philosophical intuitions.

So one of the reasons we would like to learn more about the mechanisms and processes that generate TEIJs is the hope that it might help provide a solution to the calibration problem. If we can come to understand how the intuitions are generated, then we might gain a better understanding of their reliability. As De Cruz puts it, "...the psychological underpinning of intuitions can provide a test key with which to test their validity" (2015, p. 236). Similarly, Knobe and Nichols propose that, "[f]irst we use the experimental results to develop a theory about the underlying psychological processes that generate people's intuitions; then we use our theory about the psychological processes to determine whether or not those intuitions are warranted" (2008, p. 8).

However, there is a fundamental problem with this approach. The standard way we normatively evaluate the functioning of a cognitive process or mechanism is to first evaluate the quality of the cognitive states it produces in different modes and circumstances; then, we use that assessment to evaluate the quality of the producer in those different modes and circumstances. In other words, we evaluate the normative status of the cognitive mechanisms and processes by virtue of what they produce. For instance, we determine how well a sensory system is functioning in different circumstances by examining the quality of the perceptual states it produces under those conditions. It is functioning properly when it produces normatively appropriate (e.g., accurate, sufficiently clear, timely, etc.) perceptual states. And it can often be characterized as malfunctioning (or irrational, or a flawed heuristic, etc.) when it produces states that are not normatively appropriate – states that are false or irrational or in some other way inadequate. We gain a sense of how reliable a cognitive system is in various contexts by first ascertaining the proportion of truthful or accurate states it produces in that context. And this, of course, requires us to be able to first assess whether or not a state that is produced is actually true or false.

Consequently, because the evaluation of the truthfulness or normative value of TEIJs is deeply problematic, the normative status of the cognitive processes and mechanisms that produce those intuitive judgments is also deeply problematic. Cummins' calibration problem is generally regarded as a problem with calibrating the intuitions themselves – with being able to assess the truthfulness of their contents. However, we can now see that it extends beyond our evaluation of the TEIJs. It also applies to the evaluation of the processes that produce those intuitions because we have no clear way to calibrate the functioning of these processes without first calibrating their output. In other words, what I am calling the secondary calibration problem is based upon the primary calibration problem – the fact that we lack

independent access to philosophical truths prevents us from evaluating the accuracy of philosophical intuitions. Because we cannot evaluate the accuracy of the intuitions, we can't normatively evaluate the cognitive mechanisms that generate them. So the hope of many, that we can overcome the calibration problem by learning more about the etiology of TEIJs, is deeply misguided because it ignores the fact that the normative assessment of cognitive producers usually requires a prior assessment of the cognitive product – the very thing that is missing in the case of TEIJs.

The findings of experimental philosophy exacerbate the secondary calibration problem because they reveal considerable variance in the intuitive reactions to particular vignettes and thereby raise the question of how we can know which of these different intuitive reactions are correct. But even if this variance had not been demonstrated, or even if widespread agreement had been shown, we would still have a legitimate concern that this agreement reflects nothing more than a systematic delusion. We already know that in the case of several cognitive processes, including different problem-solving strategies, people predictably and systematically generate faulty and inaccurate mental states (see, for example, Nisbett and Ross, 1980; Kahneman, Slovic and Tversky, 1982). We know that many of our cognitive faculties systematically fail to produce normatively correct intuitions, especially in particular contexts. But we know this only because we can test the mental states produced in those contexts. We can measure the lines in a Mueller-Lyer drawing and show that our visual seemings are incorrect. We can do careful statistical analysis to show that widespread intuitions about probabilities are mistaken. Using these findings, we can *then* theorize about the normative nature of the cognitive machinery that produces these intuitions. But without these findings about their output, we would have no basis for evaluating their functioning. And it is precisely this sort of finding about the status of TEIJs that we don't have. Consequently, we have no basis for evaluating the functioning of the cognitive mechanisms that generate TEIJS, and thus we cannot use such an evaluation of the functioning of those cognitive mechanisms to provide an assessment of those TEIJs.

Another way to think about this issue is in terms of Chomsky's well-known distinction between competence and performance (Chomsky, 1965). Chomsky and other linguists argue that an underlying linguistic competence involving some sort of knowledge structure, like a grammar, produces much of our linguistic output, including linguistic intuitions. However, in certain situations and contexts, our actual linguistic performance does not properly reflect the nature of this underlying competence because of performance errors, caused by influences or limitations of other important sub-systems, such as working memory or attention. Note that we can call the output in such situations performance errors because we have a method for evaluating the status of the linguistic output. We can, for example, describe the faulty intuition that sentence (A) is ungrammatical by carefully examining it and

seeing that, in fact, it properly follows the rules for multiply-center-embedded clauses, albeit in a very complicated way:

(A) A man that a woman that a child that a bird saw knows loves.

We can then speculate about the cognitive source of this intuitional error, and assign it to something like a failing in working memory. In any event, this breakdown in the production of linguistic intuition can only be ascertained once we have first determined the mistaken nature of the intuition itself. And this is only possible because we have an independent source of information (English grammar) about center-embedded clauses that reveals the flawed nature of the intuition. If we had no way to assess the accuracy of linguistic intuition, and all we had to go on was the intuition itself, then it would be very difficult to see how we could distinguish between processes and systems that contribute to linguistic competence, or the expression of linguistic competence, and processes that instead contributed to some sort of intuitional error. Any talk of systematic performance errors presupposes the ability to determine flawed output. And that ability appears to be exactly what is missing in the case of TEIJs.

One possible strategy for addressing this issue and solving the calibration problem has been proposed by Talbot (2009) and is at least suggested by De Cruz (2015).[7] The proposal is that we can learn about the reliability of processes generating intuitions on mundane matters for which we actually *do* have independent access to the truth. We can then extend what we learn to the more enigmatic TEIJs and how they are produced. As Talbot puts it:

> We can determine how intuitions work – the data they are likely to be sensitive to and the data they are likely to ignore, and what factors make them more or less accurate – by studying intuitions about non-philosophical questions we know the correct answers to. These are questions about ordinary objects, behavior, possibilities and so forth. We can compare what we learn about how intuitions do and do not work for ordinary questions with our demands on a source of evidence for philosophical questions, and calibrate our intuitions in this way (2009, p. 165).

Similarly, De Cruz suggests that if Gettier intuitions arise from folk psychological mechanisms, as suggested by Nagel (2012), then we can evaluate the accuracy of those intuitions by generalizing from the accuracy of folk psychological intuitions on more mundane, testable matters (De Cruz, 2015, p. 9).

While I regard these proposals as interesting, there are two reason why I am not very optimistic about using this strategy for evaluating philosophical intuitions and their psychological origins. The first is that a well-known source

---

7   I am grateful to an anonymous reviewer for correctly pointing out that I should address Talbot's and De Cruz's proposals.

of cognitive error is the employment of cognitive processes or mechanisms in contexts that are different from those in which they function properly. As Nisbett and Ross put it in their seminal text on cognitive errors: "We contend that people's inferential strategies are well adapted to deal with a wide range of problems, but that these same strategies become a liability when they are applied beyond that range..." (Nisbett and Ross, 1980, xii). This point applies to a variety of cognitive mechanisms and operations, including whatever psychological systems and sub-systems that generate intuitions. Just because a cognitive strategy produces reliable intuitions about some ordinary testable matter (such as, say, whether or not a physical object could be used as a hammer), this gives us no reason to think the same strategy would be reliable – or even work at all – to create true intuitions about the nature of knowledge, free action, reference or any of the other abstract matters philosophers care about. Indeed, as Machery has recently argued (Machery 2017), philosophical cases are more prone to generating intuitional errors because they typically involve judgments about bizarre scenarios with "disturbing characteristics", like highly unusual circumstances and separating properties that normally go together. Intuitions about the ordinary can't go very far in helping us calibrate philosophical intuitions precisely because the latter are almost always about the extraordinary.

The second problem with this proposal stems from the etiological diversity challenge we discussed in the last section. Talbot and De Cruz's calibration method assumes a high degree of uniformity in the production of intuitions, since the processes and mechanisms need to be relatively constant and invariable for the lessons learned about intuitions in one domain to apply to others. But as we saw in Section III, this sort of uniformity assumption is highly dubious.[8] Not only is it doubtful that the exact same cognitive machinery that generates intuitions about ordinary matters also serves to generate intuitions about philosophical cases, but it is also doubtful that the cognitive machinery that generates intuitions about particular philosophical cases in certain contexts is the same machinery that generates intuitions about other philosophical cases or even similar philosophical cases in other contexts. This proposed solution to the calibration problem requires precisely the sort of generalizations about the etiology of intuitions that the etiological diversity problem prevents.

It might be supposed that we can evaluate the reliability of a cognitive process that generates a philosophical intuition by demonstrating interference – causal influence by sub-systems that have been independently determined to be inappropriate. Through independent investigations, we might be able determine that a certain module functions to generate a particular sort of psychological output, such as an emotional response or a certain

8   I'm grateful to an anonymous reviewer for pointing out the relation between the etiological diversity problem and Talbot and De Cruz's arguments about calibrating intuitions.

attitude. Then perhaps it could be argued that this output is inappropriately influencing the TEIJ in question, and thus the process could be characterized as introducing a corrupting, error-producing factor. Yet the problem with this strategy is that it is always possible for the defenders of the intuition to argue that the module's output is not inappropriate at all, and that, in fact, the error lies in any theory that claims otherwise. Indeed, a philosophical theorist that embraces the intuition can say these findings reveal an important deep truth about the philosophical issue in question; namely, that contrary to what some might have thought, it turns out that truthful intuitions about subject X require precisely the allegedly "corrupting" sort of input. It seems we can't get away from the fact that any treatment of a cognitive process as corrupting requires the prior assumption that the output it helps generate is, in fact, in error. One theorist's evidence of performance error is another theorist's evidence of underlying competence.

A nice illustration of this point involves a well-known argument by Joshua Greene challenging deontological ethical theories as an attempt to rationalize flawed moral intuitions (Greene, 2008). Greene argues that for vignettes describing certain forms of up-close and personal violence, regions of the brain known for emotional reactions become highly activated. The amygdala, for instance, which is known to be involved in the production of strong emotions, is highly activated during footbridge versions of the Trolley case (when most subjects have the intuition it is morally wrong to personally cause a death to save many lives – e.g., by pushing a man off a bridge), but it is mostly inert when the subject is presented with switch versions of the case (where most subjects have the intuition it is morally permissible to do something remote – throw a switch – that will cause the one death to save many lives). This leads Greene to suggest that the amygdala is introducing a corrupting influence of emotion in the footbridge intuitions. Greene argues that the proper moral intuition, reflecting genuine moral reasoning, is the utilitarian response exhibited in the switch case. Although deontological theories (at least ones that argue it is always wrong to cause the death of an innocent) are supported by the footbridge intuitions, Greene argues these intuitions are not proper moral reactions because of the corrupting influence of emotion.

In response to Greene's indictment of intuitions supporting deontology and opposing utilitarianism, some authors have directly challenged Greene's claim that emotional input from brain regions like the amygdala should be regarded as corrupting to moral judgment.[9] Instead, people within the "sentimentalist" tradition of deontology (or, indeed, any form of sentimentalist moral theory) argue that the production of proper moral intuitions *requires* an emotional component. For them, the influence of the amygdala is exactly what one should expect to find when people engage in proper moral reasoning and reflection. Without any independent and compelling means by which we

---

9   See, for example, Timmons (2008).

could evaluate and characterize the footbridge intuitions as flawed, Greene has no real basis for treating an emotion-generating cognitive mechanism like the amygdala as having an undesirable influence on intuition.

Consider again Knobe's interesting account of the mechanisms generating intuitions about intentional action. An important feature of Knobe's model is that the introduction of moral judgment in the process is no longer treated as a corrupting influence, but instead as part of our underlying competence in assigning intentionality. As Knobe puts it, "[t]he chief contribution of this new model is the distinctive status it accords to moral considerations. Gone is the idea that moral considerations are 'distorting' or 'biasing' a process whose real purpose lies elsewhere. Instead, the claim is that moral considerations are playing a helpful role in people's underlying competence itself" (2008, p. 145). Here again we see how a process that some might characterize as corrupting can be turned around as a proper part of our competence. This will always be possible because without an independent assessment of the status of the intuitions produced, we have no independent arbiter of what counts as corrupting vs. proper in the etiological process.[10]

This is not to say that we can't ever properly treat an *input* to the cognitive process as corrupting or contaminating to the intuition. In fact, experimental philosophy has revealed all sorts of ways in which we can show that TEIJs can be heavily influenced by factors that, normatively, ought not to have such an influence. For example, as Nichols and Knobe (2008) note, studies by Lerner et al. (1998) reveal that subjects who have their anger aroused by reading a maddening story about a bully are more inclined to assign responsibility to agents in unrelated cases of negligence. We can plausibly say that one *ought not* to judge others as more responsible for an act just because one is in an angry state, triggered by contemplation of an unrelated scenario. There is nothing problematic about identifying so called

---

10   Nichols and Knobe (2008) consider this sort of no-arbiter problem in discussing intuitions about free will. As they note, studies reveal that subjects are more inclined to have compatibilist intuitions toward a vignette that is likely to produce strong affect (like someone murdering his family) and incompatibilist intuitions for low affect cases (like someone cheating on his taxes). The question is, which of these intuitions are "legitimate": in their terminology, there are two possible models – the "performance error model" and the "affective competence model", and it is far from clear which one is correct. They offer a clever study involving deterministic and non-deterministic universes to argue that the legitimate intuitions are the incompatibilist ones and, thus, the input of strong emotion (creating compatibilist intuitions) is corrupting. While their argument is compelling, I see no reason why an equally clever defender of compatibilism couldn't counter that affect is essential for proper intuitive judgments about free will, and that without it other corrupting influences (such as, say, tacit beliefs about agent causation or some such) bias the process when there is consideration of a deterministic universe. A similar concern is raised by Alexander, Mallon and Weinberg (2014), who point out that, "a proponent of an affective competence model could suggest that people's answers in the high-affect cases are fine, but that some other mechanism interferes with people's judgments in the low-affect cases..." (p. 41).

framing effects – ways in which it can be experimentally demonstrated that an intuitive reaction is influenced by factors that ought not to make a significant difference, like having watched a funny skit from Saturday Night Live (Valdesolo and DeSteno, 2006), or having the paper presenting the vignette sprayed with Lysol (Tobia et al., 2013). We can, in other words, employ a prior understanding of what should and should not be relevant to the determination of certain philosophical matters to treat specific inputs as contaminating if they influence a judgment about those matters.

It is important to note, however, that many of these revelations of contaminating or corrupting factors are not based upon the normative assessment of the cognitive machinery that produces the TEIJs. In accounts such as Lerner et al.'s, there is relatively little known about the actual cognitive machinery generating the judgment. Instead, these evaluations are based upon prior beliefs about the sort of factors that are relevant to the type of TEIJ generated. We say an intuition about the responsibility of a suspect should not be influenced by being in a state of anger caused by something completely unrelated to the crime because we have beliefs and attitudes about guilt that make such an influence illegitimate. The identification of framing effects is not an assessment of the output (TEIJ) based upon an assessment of the machinery that produces it. It is an assessment of the output based upon an assessment of the *input* to the machinery that produced it – that such input ought to be treated as irrelevant. Further understanding of the cognitive etiology is unlikely to have much bearing on that assessment.

Couldn't we use the assessment of corrupting input to impugn certain mechanisms or sub-systems, and then use the activation or engagement of such mechanisms to impugn the TEIJ? We might, for instance, show that a certain brain region becomes highly activated whenever an unusually disgusting scenario is presented, and then argue that it is something like a 'disgust affect generator'. Then, if such a mechanism is shown to contribute to a TEIJ in which a feeling of disgust should be irrelevant, we could use the influence of such a mechanism to challenge the epistemic status of the TEIJ. I don't want to claim that such a scenario is impossible, so I'll allow that this is perhaps a way in which the etiological project could show that a TEIJ is faulty. However, I suspect that this would really be another case where the inappropriate input would be providing the evidence of corruption, not an evaluation of the mechanism. For suppose that the mechanism was activated without the relevant input. One might then think that this would be an incontrovertible case where we could legitimately say a cognitive sub-system is corrupting the intuition-generation process (and not the input itself). However, it strikes me as more likely that one of two things would occur: Friends of the intuition would either argue that since the mechanism is active without being prompted by disgusting input, the mechanism actually *isn't* simply 'disgust affect generator' after all. *Or* the case reveals that, surprisingly enough, a little disgust is needed for the appropriate intuitive response! Either way, the case would be far from incontrovertible.

To sum up this section, Cummins' original calibration claim is about the psychological states themselves, the intuitings (the TEIJs). His claim is that we have no way to calibrate them, no independent way to assess the truthfulness of their contents. In reply, it has been suggested that perhaps we *can* calibrate them by closely examining the cognitive mechanisms that produce them and then use our evaluation of the mechanism's reliability to assess the TEIJs. And my rebuttal to that claim is that this won't work because, normally, we first need an assessment of the *output* of cognitive mechanisms to determine the mechanisms' reliability; we can't, in other words, put the mechanisms-assessment cart before the output-assessment horse. Cummins' calibration problem cannot be resolved by looking at the cognitive mechanisms and processes that produce TEIJs because the calibration problem extends to our evaluation of those cognitive processes and mechanisms themselves. Normative assessments of cognitive processes typically require normative assessment of their outputs, and that is precisely what is missing given the original calibration problem. While perhaps there will be some limited ways in which a cognitive process can be impugned as corrupting, I suspect these will be rare, given that friends of the intuition can always claim that the allegedly corrupting element is really part of our underlying competence. Finally, it should be noted that the most of these strategies are ones that could be used to show that an intuition-producing process is flawed or in error. I have no idea how we could demonstrate that an etiological process is reliable and produces appropriate intuitions. Given that one of the main goals of the etiological program is to show if and when intuitions are accurate, this is a major limitation.

V. Conclusion

Assuming the arguments presented above are sound, what should be the way forward with regard to our investigation of the etiology of philosophical intuitions? While I have offered a somewhat pessimistic outlook about what such an investigation is likely to yield, I certainly do not want to imply that I think investigating the psychological source of philosophical intuitions should stop or is a worthless enterprise. Quite the contrary – I believe that there is a great deal to be learned that will be valuable, and that although we may be unable to answer all of the original questions we began with, we can still address a number of important ones.

In light of the etiological diversity problem, how should we proceed? Although I believe we need to substantially temper or even deflate our expectations for the etiological project, the good news is that we can and should continue doing much of what we saw in Section III is already being done by many researchers; namely, rather than looking for a psychological account of *the* faculty of intuition, instead investigating how particular individual TEIJs

are created in particular situations and with particular populations. Rather than trying to find an account that explains how epistemological intuitions or metaphysical intuitions are generated, we should instead focus on explaining individual reactions to specific thought experiments, trying to construct a plausible model tailored for each one. Perhaps in certain cases the application of some sort of heuristic makes the most sense, whereas in others it might involve the application of a specific type of concept, whereas with others maybe some other sort of complex process is involved. This means there won't be *an* etiological project, but many. While investigating the cognitive source of each individual sort of TEIJ may sound daunting, the list of important and impactful intuition-producing thought experiments in philosophy is not terribly long. And as we saw, much of this sort of work is already underway in the cognitive program of experimental philosophy, especially in the area of moral psychology.

What about the secondary calibration problem? As I noted in section IV, there may be some clever ways in which the reliability of a TEIJ can be called into doubt by focusing on cognitive machinery producing it. But I am inclined to think this will be the exception and not the norm. And while the unreliability of intuitions can occasionally be shown, I am much less optimistic about the prospects of demonstrating the trustworthiness of intuitions in this way. Thus, this problem adds to a growing list of considerations that have motivated several authors to challenge the import we typically assign to philosophical intuitions prompted by thought experiments.[11] Given these considerations, it makes sense to step back and ask what we want an account of some philosophical notion to do for us. Traditionally, we want that account to accurately capture some sort of mind-independent deep philosophical truth, and the hope was that intuition would guide us to such an account. However, if there is no way to demonstrate how and when the cognitive machinery producing TEIJs succeeds in relaying or transmitting these philosophical truths, then perhaps it is time to reconsider how we assess our philosophical theories.

One sensible option is to recognize that philosophical notions like knowledge, justice, freedom, responsibility, reference, moral goodness and the like all play very important roles in different aspects of our lives, including social institutions, legal systems, politics, science, normativity and so on. Instead of treating TEIJs as something like observational evidence, and thereby treating them as of central importance in determining which philosophical accounts we ought to embrace, we can and probably should diminish their significance and allow other, more pragmatic considerations take center stage. We should favor theories not because they are sanctioned by intuition, but because they put forward philosophical principles and concepts that do the sort of work we need them to do. Take, for instance,

---

11   See, for example, Machery (2017), Weatherson (2003) and Weinberg, Nichols and Stich (2001).

our conception of knowledge and the Gettier intuitions that challenge the justified, true belief account. Given that the justified true belief account is straightforward, easy to use, and works quite well in distinguishing knowledge from non-knowledge in nearly every important and realistic setting (such as in distinguishing science from non-science), and given that we have no way of ascertaining the accuracy of the Gettier intuitions, a reasonable response would be to simply ignore them (Weatherson, 2003). This is just one example of how, although the etiological project cannot inform us about a faculty of intuition (because there is none), nor conclusively show that intuitions are produced in a reliable manner, it can nevertheless motivate a sensible path forward by diminishing our reliance upon intuitions. Indeed, the problems that plague the etiological project may eventually prove to be benefits insofar as they help promote a more reasonable and pragmatic outlook on theory development in philosophy.

References

Alexander, J., 2012, *Experimental Philosophy: An Introduction*, Malden, MA, Polity Press.

Alexander, J., Mallon, R. and Weinberg, J., 2014, "Accentuate the Negative", in J. Knobe and S. Nichols (eds.), *Experimental Philosophy, Volume 2*, Oxford, Oxford University Press.

Alexander, J. and Weinberg, J., 2014, "The 'Unreliability' of Epistemic Intuitions", in E. Machery and E. O'Neill (eds.), *Current Controversies in Experimental Philosophy*, New York, Taylor and Francis, pp. 128–145.

Allman, J. and Woodward, J., 2008, "What are Moral Intuitions and why Should We Care About Them? A Neurobiological Perspective", *Philosophical Issues*, 18, pp. 164–185.

Arico, A., Fiala, B., Goldberg, R., and Nichols, S., 2011, "The Folk Psychology of Consciousness", *Mind and Language*, Vol. 26:3, pp. 327–352.

Barsalou, L. W., Huttenlocher, J., & Lamberts, K., 1998, "Basing Categorization on Individuals and Events. *Cognitive Psychology*, 36, pp. 203–272.

Bealer, G.,1998, "Intuition and the Autonomy of Philosophy," in M. DePaul and W. Ramsey, *Rethinking Intuition*, Lanham, MD, Roman and Littlefield, pp. 201–240.

Chomsky, N., 1965, *Aspects of the Theory of Syntax*, Cambridge, MA, MIT Press.

Cummins, R., 1998, "Reflections on Reflective Equilibrium," in M. DePaul and W. Ramsey, *Rethinking Intuition*, Lanham, MD, Roman and Littlefield, pp. 113–128.

De Cruz, H., 2015, "Where Philosophical Intuitions Come From", *Australasian Journal of Philosophy*, Vol. 93, No. 2, pp. 233–249.

Earlenbaugh, J. and B. Molyneux, 2009, "Intuitions are Inclinations to Believe," *Philosophical Studies*, 145, pp. 89–109.

Evans, J. S. B. T., 2010, *Thinking Twice: Two Minds in One Brain*, Oxford: Oxford University Press.

Evans, J. S. B. T. and Frankish, K., 2009, *In Two Minds: Dual Processes and Beyond*, Oxford: Oxford University Press.

Evans, J. S. B. T. and Over, D., 1996, *Rationality and Reasoning,* Hove, UK: Psychology Press.

Fischer, E. and Collins, J., 2015, "Rationalism and Naturalism in the Age of Experimental Philosophy", in E. Fischer and J. Collins (eds.), *Experimental Philosophy, Rationalism and Naturalism: Rethinking Philosophical Method*, London, Routledge Press, pp. 3–33.

Gigerenzer, G., Todd, P. M., & the ABC Research Group, 1999, *Simple Heuristics that Make us Smart*, New York: Oxford University Press.

Goldman, A., 2007, "Philosophical Intuitions: Their Target, Their Source, and Their Epistemic Status," *Grazer Philosophische Studien*, 74, pp. 1–26.

Greene, J., 2008, "The Secret Joke of Kant's Soul", in W. Sinnott-Armstrong (ed.) *Moral Psychology, Volume 3: The Neuroscience of Morality: Emotion, Brain Disorders, and Development*, Cambridge, MA, MIT Press, pp. 35–80.

Hales, S. D., 2012, "The Faculty of Intuition", *Analytic Philosophy*, Vol. 53:2, pp. 180–207.

Hauser, M., Young, L., and Cushman, F., 2008, "Reviving Rawls's Linguistic Analogy: Operative Principles and the Causal Structure of Moral Actions", in W. Sinnott-Armstrong (ed.), *Moral Psychology, Volume 2 – The Cognitive Science of Morality: Intuition and Diversity*, Cambridge, MA, MIT Press, pp. 107–143.

Kahneman, D. and Frederick, S., 2005, "A Model of Heuristic Judgment", in K. Holyoak and R. Morrison (eds.), *The Cambridge Handbook of Thinking and Reasoning*, New York: Cambridge University Press, pp. 267–293.

Kahneman, D., Slovic, P. and Tversky, A. (eds.) 1982, *Judgment Under Uncertainty: Heuristics and Biases,* Camrbidge, Cambridge University Press.

Keil, F. C., 1989, *Concepts, Kinds, and Cognitive Development*, Cambridge: MIT Press.

Knobe, J., 2004, "Intention, Intentional Action and Moral Considerations", *Analysis*, 64, pp. 190–193.

Knobe, J., 2008, "The Concept of Intentional Action", in J. Knobe and S. Nichols (ed.) *Experimental Philosophy*, Oxford, Oxford University Press, pp. 129–147.

Knobe, J. and Nichols, S., 2008, "An Experimental Philosophy Manifesto", in J. Knobe and S. Nichols (ed.) *Experimental Philosophy*, Oxford, Oxford University Press, pp. 3–16.

Knobe, J. and Prinze, J., 2008, "Intuitions About Consciousness: Experimental Studies", *Phenomenology and Cognitive Sciences* 7, pp. 67–85.

Lerner, J., Goldberg, J., and Tetlock, P., 1998, "Sober Second Thought: The Effects of Accountability, Anger, and Authoritarianism on Attributions of Responsibility", *Personality and Social Psychology*, Bulletin 24.

Lewis, D., 1983, *Philosophical Papers: Volume I*, New York: Oxford University Press, pp. x.

Lycan, W., 1988, *Judgment and Justification*. Cambridge: Cambridge University Press.

Machery, E., 2009, *Doing Without Concepts*, New York, Oxford University Press.

Machery, E., 2017, *Philosophy Within its Proper Bounds*, Oxford, Oxford University Press.

McCauley, R., 2011, *Why Religion is Natural and Science is Not*, Oxford, Oxford University Press.

Murphy, G., 2002, *The Big Book of Concepts*, Cambridge, MA, MIT Press.

Nado, J., 2013, "Why Intuition"? *Philosophy and Phenomenological Research,* 86, pp. 15–41.

Nagel, J., 2012, "Intuitions and Experiments: A Defense of the Case Method in Epistemology", *Philosophy and Phenomenological Research*, Vol. 85: 3, pp. 495–527.

Nichols, S. and Knobe, J., 2008, "Moral Responsibility and Determinism: The Cognitive Science of Folk Intuitions", in J. Knobe and S. Nichols (ed.) *Experimental Philosophy*, Oxford, Oxford University Press, pp. 105–126.

Pust, J., 2017, "Intuition", *The Stanford Encyclopedia of Philosophy* (Summer 2017 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2017/entries/intuition/>.

Rosch, E., 1975, "Cognitive Representation of Semantic Categories", *Journal of Experimental Psychology*, v. 104, pp. 192–233.

Scholl, B. 2007, "Object Persistence in Philosophy and Psychology", *Mind and Language* 22, pp. 563–591.

Shieber, J., 2012, "A Partial Defense of Intuition on Naturalist Grounds," *Synthese*, 187: 2, pp. 321–341.

Sinnnot-Armstrong, W., Young, L. and Cushman, F., 2010, "Moral Intuitions" in J. Doris (ed.) *The Moral Psychology Handbook*, Oxford, Oxford University Press, pp. 246–272.

Smith, E. and Medin, D., 1981, *Categories and Concepts*, Boston, Harvard University Press.

Sosa, E., 1998, "Minimal Intuition," in M. DePaul and W. Ramsey, *Rethinking Intuition*, Lanham, MD, Roman and Littlefield, pp. 257–269.

Sytsma, J. and Livengood, J., 2016, *The Theory and Practice of Experimental Philosophy*, Peterborough, Broadview Press.

Sytsma, J. and Machery, E., 2009, "How to Study Folk Intuitions about Phenomenal Consciousness", *Philosophical Psychology* 22 (1), pp. 21–35,

Talbot, B., 2009, "Psychology and the Use of Intuitions in Philosophy", *Studia Philosphica Estonia*, 2.2, pp. 157–176.

Timmons, M., 2008, "Toward a Sentimentalist Deontology", in W. Sinnott-Armstrong (ed.) *Moral Psychology, Volume 3: The Neuroscience of Morality: Emotion, Brain Disorders, and Development*, Cambridge, MA, MIT Press, pp. 93–104.

Tobia, K., Chapman, G., and Stich, S., 2013, "Cleanliness is Next to Morality, Even For Philosophers", *Journal of Consciousness Studies*, 20, pp. 195–204.

Valdeso, P. and DeSteno, D., 2006, "Manipulations of Emotional Context Shape Moral Judgment", *Psychological Science*, 17, pp. 476–477.

Weinberg, J. and Alexander, J., 2014, "The Challenge of Sticking With Intuitions Through Thick and Thin", in A. Booth and D. Rowbottom (eds.), *Intuitions*, Oxford, Oxford University Press, pp. 187–212.

Weinberg, J., Nichols, S., and Stich, S., 2001, "Normativity and Epistemic Intuitions", *Philosophical Topics*, 29, pp. 429–460.

Weiskopf, D., 2009, "The Plurality of Concepts", *Synthese*, 169, pp. 145–173.

Wetherson, B., 2003, "What Good are Counterexamples?", *Philosophical Studies*, Vol. 115–1, pp. 1–31.