Accentuate the Negative

JOSHUA ALEXANDER, RONALD MALLON, AND JONATHAN M. WEINBERG[1]

(Forthcoming in the *Review of Philosophy and Psychology*)

1 Introduction

There are a number of different programs that fall under the umbrella of "experimental philosophy", and our interest here is to drive a wedge of contention between two of them. These two programs concern traditional analytic philosophy's practice of appealing to philosophical intuitions either as evidence for (or against) philosophical claims, or as data both about the nature of our folk philosophical concepts and judgments and about the nature of the domains in which we make those judgments.[2] According to what is sometimes called experimental philosophy's "negative program", experimental philosophy challenges the usefulness of this practice in achieving justified beliefs.[3] According to experimental philosophy's "positive program", experimental philosophy is (at least an indispensable

[1] Each author contributed equally to the project; authors are listed alphabetically. We would like to thank the participants in the 2008 Society for Philosophy and Psychology Workshop on Experimental Philosophy; the members of the Experimental Epistemology Laboratory (EEL) at Indiana University, especially Cameron Buckner and Chad Gonnerman; and two anonymous referees for this journal for valuable feedback on earlier drafts of this paper.

[2] In this paper, we will take as our target philosophical intuitions as they are standardly conceived of in current practice. According to this conception, philosophical intuitions are propositional attitudes generated in response to hypothetical cases in philosophy which are "minimally foundational" (a person may appeal to them as evidence without having to provide evidence for them), non-inferential, and fallible. We also think that most of these arguments will go through *mutatis mutandis* for other conceptions of such judgments and their place in philosophical methodology (see, for example, Williamson 2004, 2005, 2007; Alexander and Weinberg 2007).

[3] The terms "negative" program and "positive" program are now in common use. We are unsure of their origin though they may have been introduced by Farid Masrour.

part of) the proper methodology for this practice.[4] In this paper, we contend that the practice of appealing to intuitions, even as modified by the positive program, still faces significant challenges from the results of the negative program.

We identify four different positive programs: *direct extramentalism*, *semantic mentalism*, *conceptual mentalism*, and *mechanist mentalism*. Each of these positive programs share at least two commitments: that intuitions are a trustworthy source of evidence or data; and that intuitions about a particular hypothetical case will, by and large, be stable and shared. However, recent empirical work conducted by philosophers and psychologists has revealed significant (and surprising) inter- and intra-personal intuitional instability. As such, positive programs face the challenge of accommodating the results of negative experimental philosophy. Some positive programs (namely, the various forms of mentalism) seem, at first glance, to be well-suited to meet this challenge. But we argue that these forms have their own problems, and so conclude that positive experimental philosophy seems to be *almost* as challenged by the results of negative experimental philosophy as is more traditional armchair analytic philosophy.

2 The Positive Program

In order to canvass the problems for the positive program, we begin by recognizing that there are a variety of positive programs extant in this still-young literature. These programs share – both with one another and with more traditional analytic philosophical programs – the view that intuitions provide an important source of evidence and data for philosophy. What distinguishes the various positive programs from more traditional analytic philosophical programs is the way in which we are supposed to go about gathering this evidence or data. According to proponents of more traditional philosophical investigation, we can determine what intuitions are (or would be) generated in response to particular cases simply by determining what our own intuitions are about those cases (e.g., Jackson 1998). Assuming that our own intuitions are appropriately representative, we need nothing more than our own intuitions about particular cases in order to determine what intuitions people would (or should) have about those cases. Proponents of positive experimental philosophy think that we would do better to actually empirically ascertain – typically employing survey methods – what intuitions people have about those cases.

[4] For additional discussions of experimental philosophy's positive and negative programs, see Alexander and Weinberg (2007); Kauppinen, (2007); Nadelhoffer and Nahmias (2007); and Weinberg (2007).

While the positive programs share common views about the nature of philosophical evidence and appropriate methods of evidence collection, they can be distinguished according to what they take the immediate philosophical payoff of experimental philosophy to be. The most fundamental question for any program of positive experimental philosophy is whether it aspires to knowledge of "in-the-head psychological entities" or "outside-the-head nonpsychological entities" – positions Alvin Goldman and Joel Pust have called *mentalist* and *extramentalist* (Goldman and Pust 1998, pp. 183-4). Goldman and Pust introduced the distinction in terms of possible rationales for armchair deployments of intuition, but it applies equally well here: if the experimental results of positive experimental philosophy are meant to tell us something of philosophical import, what type of thing is it supposed to be? The answer one gives to this question will determine the most basic theoretical burdens that it must shoulder. We will consider four possible answers to that question here, corresponding to four different philosophical projects. We do not claim that this list is necessarily exhaustive, though we do think it covers almost all extant forms of positive experimental philosophy. [5] The four projects are direct extramentalism; conceptual mentalism; semantic mentalism; and mechanist mentalism.[6]

Direct extramentalist projects are those that draw conclusions about nonmental entities from premises that include empirical claims about folk intuitions or judgments but do not include premises about human psychology arrived at by via those empirical claims. Direct extramentalists take a proposition's status as intuitive to be direct evidence for the truth of that proposition, even if perhaps not conclusive evidence. For example, they might take it that philosophical positions that are intuitive to a large majority of ordinary people and that are not matters of technical expertise should be given a significant default positive epistemic status. So, if most folks are intuitive compatibilists, then that is evidence for compatabilism, and perhaps incompatibilists should have the burden of proof in debates over free will – and *vice versa*, *mutatis mutandis*. A practitioner of this version of direct extramentalist positive experimental philosophy might hope that experimental psychological methods could then uncover which of those views has that argumentative burden (Nahmias, Morris, Nadelhoffer, and Turner 2006).

But one may be concerned (as Goldman and Pust are) that direct extra-

[5] One notable exception is the recent work of Stotz and Griffiths (2004). They document the varying intuitions of specialist populations regarding the concept of the gene, and they have a good reason for restricting their populations of interest. Moreover, there are a number of interesting examples of experimental philosophy that do not particularly concern intuitions (see, e.g., Nichols 2002, Schwitzgebel forthcoming), and that our arguments do not target.

[6] See Nadelhoffer and Nahmias (2007) for an earlier but different elucidation of forms of positive experimental philosophy.

mentalism allows too ambitious an evidential role for intuitions. At best, one might argue, intuitions reflect facts about our minds, and it is only in virtue of our philosophical interest in these mental facts that intuitions can play their methodological role. The current positive experimental philosophy literature reveals a number projects that fall under this "mentalist" rubric albeit in a range of different ways.

Conceptualist mentalist projects take an interest in what actual conceptual structure is instantiated in people's heads, for various concepts of philosophical interest, such as INTENTIONAL (Knobe 2003) or INNATE (Machery, Griffiths, and Lindquist, forthcoming). *Semantic mentalist* projects are concerned to identify the meanings of our terms or concepts. For example, following the program of philosophical analysis of folk concepts that runs through David Lewis (1970, 1972) and Frank Jackson (1998), some positive experimental philosophers have insisted that the evidence that some proposition is a folk platitude be empirically supported – or at least empirically scrutinized (see, e.g., Glasgow 2008, Ulatowski 2008). Finally, *mechanist mentalist projects* aim to understand the psychological structures and processes involved in our making judgments in a domain of psychological interest. Can our folk psychology be understood in primarily prediction-and-explanation terms, or is it deeply entwined with our moral and evaluative cognition as well? (Knobe 2003, 2007b) To what extent do affect and rules contribute to the difference between normative evaluations that are moral and those that are not? (Nichols and Mallon 2006; Mallon and Nichols, forthcoming).

Mentalist projects take knowledge of the mental to be not just the immediate philosophical payoff of surveying the intuitions of various subjects; a further subdivision is possible, according to whether that knowledge of the mental is the main philosophical payoff of the project, or whether instead such claims serve primarily as intermediate steps in a further philosophical argument. Joshua Knobe (2007a), for example, has argued persuasively for the philosophical legitimacy of the former sort of project. But some positive experimental philosophy offers the promise of speaking to philosophical questions beyond the mind itself; for example, a theorist may look to take an area of philosophy in which we have had conflicting intuitions, and deploy a psychological theory of those intuitions' production in order to help referee which should be trusted, and which merely explained away (e.g., Greene 2003; Nichols 2006).

3 The Pitfalls of Positive Experimental Philosophy

A The Empirical Challenge from Negative Experimental Philosophy

As we noted earlier, positive experimental philosophy shares with traditional armchair philosophy the commitment that intuitions about X are a trustworthy source of evidence or data for philosophical theorizing about X (or at least about "X" or the concept of X); and that intuitions about a particular hypothetical case will, by and large, be shared, at least by "the folk". But some recent empirical work conducted by philosophers and psychologists gives us reason to worry that philosophical intuitions might be neither trustworthy nor shared. They suggest that some particularly prominent, and commonly appealed to, philosophical intuitions are sensitive to facts about who is considering the hypothetical case[7], the presence or absence of certain kinds of content (e.g. abstract vs. concrete; affectively neutral vs. affectively engaging)[8], or the context in which the hypothetical case is being considered.[9] This sensitivity is problematic because such facts have not traditionally been thought to be relevant to the truth or falsity of the claims for which philosophical intuitions are supposed to provide evidence or data. Additionally, when these studies are coupled with our inability to either explain what it is about any of these intuitions that make them problematically sensitive or predict which other intuitions may or may not be problematically sensitive, they challenge the trustworthiness, not just of the class of intuitions that have so far been studied, but of the whole class of intuitions typically appealed to in philosophical discourse (Alexander and Weinberg 2007; Weinberg 2007).

In addition to calling into question the trustworthiness of philosophers' typical appeals to intuitions, these recent empirical studies also call into question whether there is, in fact, something like a shared intuition about a particular hypothetical case that can be appealed to either as evidence or data. These studies show that particular hypothetical cases can give rise to a number of different intuitions, thereby calling into question any claims as to what *the* folk intuitions are – a significant problem for positive programs, each of which views getting at *the* folk intuitions to be either a significant philosophical insight in its own right or a necessary step towards achieving a significant philosophical insight. It also raises the question of how we should proceed when confronted with conflicting intuitions. At a bare minimum, anyone who wants to select one from among those intuitions that are generated in response to a given hypothetical case needs to explain why the other intuitions should be discounted. The trouble is that determining just what to do when confronted with conflicting evidence or data is not especially straightforward, as the growing literature in the epistemology of disagreement demonstrates (see, e.g., Christensen 2007; Elga 2006;

[7] Weinberg, Nichols, and Stich (2001) and Machery, Mallon, Nichols, and Stich (2004).
[8] Nichols and Knobe (2007) and Pizzaro *et al.* (Manuscript).
[9] Swain, Alexander, and Weinberg (2008) and Petrinovich and O'Neill (1996).

Feldman 2006; Feldman and Warfield 2007; Kelly 2005, 2007, and 2008; and White 2005).

These findings thus pose a significant challenge to positive experimental philosophy, inasmuch as positive experimental philosophy attempts to deploy intuitions as evidence or as data. In fact, we think that positive experimental philosophy may be *almost* as imperiled by negative experimental philosophy as is more traditional armchair analytic philosophy. Positive experimental philosophers need a way to accommodate the kinds of inter- and intra-personal differences discovered by negative experimental philosophers. Direct extramentalist projects appear lethally imperiled: they involve inferences from premises of the form "it is intuitive that P", but such premises now seem ill-formed, without specifying *to whom* it is intuitive, and *under what circumstances*.[10] However, the various mentalist positive programs may seem, at first glance, to be well situated to pull off just such a needed accommodation. But, as we shall see, these mentalist programs face equally challenging problems.

B Conceptualist and Semantic Mentalism and the Machery-Quine Problem

Conceptualist approaches to positive experimental philosophy have proved popular, and are perhaps the most common sort of experimental philosophy today. Although conceptualists have not been motivated by the worries we reviewed in the previous section, conceptualist mentalism seems to hold out the *prima facie* promise of some resources that would make it better able to withstand the challenge from negative experimental philosophy. First, conceptualist mentalism allows for a modicum of relativization, which may go some way towards defusing the threat of cross-group differences in intuitions – if Asian and Western subjects have different intuitions, then perhaps they just have different concepts (though see Mallon, Machery, Nichols, and Stich forthcoming). Relatedly, one may hope under conceptualist mentalism to be able to disregard some variation and instability of intuitions as mere noise, not reflective of the underlying concept. The two moves are related, as the first one is only possible if the second one can enable us to

[10] Perhaps some form of relativism or contextualism could be attempted. Different relativizations might have different degrees of plausibility (for example, see Glasgow 2008 on relativism and the concept of race). While such moves may be appropriate in some instances, we suspect that they will not prove generally attractive. For a discussion of why epistemic contextualism might not be particularly helpful, see Swain, Alexander, and Weinberg 2008. For a discussion of relativizing intuitions about reference, see Mallon, Machery, Nichols and Stich forthcoming.

distinguish conceptually-based differences in intuition from non-conceptually-based differences.

Conceptualist mentalism thus relies on the idea that we can use empirical evidence to establish what is and what is not constitutive of a given concept of philosophical interest. But it is, in fact, far from clear how to do so – a point made in the context of experimental philosophy by Edouard Machery (2008) in his discussion of the debate over the so-called "Knobe-effect" or "side-effect effect." In probably the most famous finding of experimental philosophy, Joshua Knobe (2003a) showed that whether or not a foreseen side effect is judged to be intentional is influenced by whether or not the side effect is bad. Knobe presented subjects with two versions of the following vignette:

> **Harm Condition**
>
> The vice-president of a company went to the chairman of the board and said, 'We are thinking of starting a new program. It will help us increase profits, *but it will also harm the environment*.'
>
> The chairman of the board answered, 'I don't care at all about *harming* the environment. I just want to make as much profit as I can. Let's start the new program.'
>
> They started the new program. Sure enough, the environment was *harmed*.

In the second, Help Condition, the vignette was the same, except the word "harm" was replaced with "help." In each case, subjects were then asked whether or not the chairman harmed/helped the environment intentionally. But the conditions produced sharply divergent results. Most subjects in the Harm Condition (82%) said the chairman harmed the environment intentionally, while most in the Help Condition said the chairman did not help the environment intentionally. Knobe (2003b) concluded that this asymmetry was not a mistake by subjects, but rather reflected the structure of the concept INTENTIONALLY. Other commentators have disputed this, alleging that the effect emerges from considerations extrinsic to the concept, for example, a desire to blame the perpetrator of foreseen harm (see, e.g., Nadelhoffer 2004a, 2004b; Adams and Steadman 2004a). Machery, however, rightly pointed out is that the debate seems to hinge upon the appropriate individuation of the concept *intentionally,* and that there is simply no way to resolve this debate absent some specific idea as to how to individuate concepts. And positive program philosophers have not taken on the difficult task of defending such an idea and figuring out how to implement it in their methods.

Moreover, the most common criterion for philosophers for individuating a concept is semantic. That is, when faced with the question of whether an inference involving a bit of mental syntax, e.g. INTENTIONAL, does or does not figure as part of the concept, philosophers ask if the inference fig-

ures in constituting the meaning of "intentional" – and here conceptual mentalism gives way to semantic mentalism. But as Quine noted in "Two Dogmas," it is not at all clear in virtue of what facts we can adjudicate disputes about meaning (Quine 1951).

Thus, this one problem seems to us to bedevil both conceptualist mentalist and semantic mentalist approaches to experimental philosophy: semantic mentalists owe a way of distinguishing meaning-constituting facts from non-meaning-constituting ones; conceptualist mentalists owe either that or some other way of individuating conceptual structures from causally interwoven, but external, structures.

One option for either project might be to hope that sophisticated psychological inquiry might solve this problem. For example, "theory" theorist psychologists aim to discern deep principles of "core knowledge" or commitments that might pull apart confounding factors, perhaps revealing the semantic structure of ordinary concepts. For instance, there is now a widespread literature discussing folk essentialist construals of various natural kind concepts (e.g., Gelman 2003), and one might view these studies as revealing necessary or sufficient conditions for membership in the kind. Given important psychological research programs such as these it may seem presumptuous, and downright unQuinean, to try to use a philosopher's armchair argument to attack a scientist's way of arguing.

We confess to not seeing how even such sophisticated psychological inquiries distinguish, on empirical grounds, what properly belongs to a concept or a meaning, and what does not. But in large part we take this to be so because we also take it that one should not simply assume that the psychologists' projects and the philosophers' projects are the same. If, instead of trying to elucidate word meanings, psychologists are typically just trying to map out what are the psychological structures and processes that implement our abilities to categorize, with no attendant commitments to any aspect of those structures being conceptually discrete or meaning-constitutive, then this "Machery-Quine" problem just doesn't arise for them. Psychologists studying concepts do not seem to us to have been particularly interested in refereeing issues of which subtle differences in categorization represent real conceptual differences. Lexical semanticists, too, have seemed mostly interested in the behavior of whole classes of words (such as unaccusative vs. unergative verbs), and have not felt the need to develop the resources that positive experimental philosophers would need to, say, discern whether or when disagreement over Gettier cases would mean that East Asians and Westerners mean something different by "knows".[11] If we are right, then it does not follow from the success of empirical psychological

[11]See, e.g., Johnson (2008). Note though that lexical semanticists might offer some resources that philosophers would find useful if the factivity of "knows" or other epistemologically-interested verbs was under discussion.

work that psychologists have a solution to either concept or meaning individuation to offer philosophers (cf., Machery 2009, Chapters 1-2).

It may be objected that that there are at least *some* instances in which scientists have employed machinery meant to distinguish properly conceptual or semantic from other psychological sources of behavior. Perhaps the most famous such machinery (though perhaps the only such machinery) for philosophers is the semantics/pragmatics distinction. And we do not wish to reject that distinction, or the wealth of tests (cancellability, etc.) that are available for making it in real scientific practice. And, furthermore, it is a distinction that has even been successfully appealed to in some debates within positive experimental philosophy (e.g., Adams and Steadman 2004a, 2004b; Knobe 2004; and Nichols and Ulatowski 2007). Nonetheless, we are not sanguine about the prospects of using this particular piece of theoretical machinery to handle the challenge advanced by negative experimental philosophy; the particular patterns of variability and instability in the negative experimental philosophy findings just do not look like the sorts of patterns one would expect, were they simply a matter of pragmatics. Nonetheless, we would certainly welcome attempts by positive experimental philosophers to do so, as a considered step in the right direction. And the richness of the theoretical background of the semantics/pragmatics distinction, with its attendant arsenal of tests for its proper deployment, just reveals how much really is required to even begin to develop tools that would begin to help answer this question of what experimental findings are truly semantic or conceptual.

Of course, as we mentioned above, a conceptualist mentalist may abandon semantic mentalism as a way of individuating conceptual structures, and offer some other, non-semantic criteria for distinguishing what does and does not properly belong to the concept or other relevant mental structure. Where meanings fail, perhaps appeal to the competence/performance distinction for, or the proper domain of, the relevant mental structure will do. We consider these in turn.

C Competence, Performance, Marr, and The Limits of Surveys

Even as negative experimental philosophy has frequently demonstrated unexpected and unwanted variation in people's intuitions, this observed variation in intuition would no longer pose a problem if we possessed a means for discerning epistemic wheat from chaff. Shaun Nichols and Joshua Knobe (2007) have attempted to do just that, with regard to divergent intuitions concerning free will and determinism, by trying to argue that some of

the observed variation is a matter of *performance errors* in one of the studied conditions.

To see how this might work, consider an example from linguistics. Linguists use intuitions about grammaticality as data to construct the grammar of a natural language, but they distinguish between the *competence* involved in producing judgments from the factors that influence *performance.* As Robert Cummins puts it: "competence is ideal... performance, that is, the performance that the system would exhibit but for resource limitations, physical breakdown, and interference from other processes" (Cummins 1996, p. 44). So, in one famous instance, subjects find multiply-center-embedded English sentences like

The man the boy the woman saw heard left.

to be ungrammatical, but linguistic theory says that they comply with the syntax with which we work (viz., they are grammatical according to our competence). The apparent ungrammaticality of these sentences is often explained away in terms of limits on working memory in the parser (e.g., Marcus 1980). Thus, we take all the evidence we have and construct a model of the cognitive mechanisms that operate to produce judgments in a domain, and we determine the borders of a folk domain by looking at the mechanism in the model that produces the paradigmatic judgments we are concerned with. The workings of that mechanism determine the competence of the subject within that folk domain. Judgments that are influenced by factors outside that mechanism represent performance errors.

Although we are not troubled by many of the uses of the competence/performance distinction throughout cognitive science,[12] we do not think that it is a distinction that can – yet – do the work that some positive experimental philosophy practitioners have hoped to have it do. Simply put, experimental philosophy currently lacks the experimental and theoretical resources to make a good use of that distinction for its purposes.

First, most experimental philosophy has been (and continues to be) reliant on survey methods. Subjects are given a questionnaire, and their judgments are elicited regarding some range of scenarios, with the experimenters typically manipulating the substance of the scenarios but also possibly their order or other contextual elements. Such methods can generate a set of extensional and typically distributional data: given scenario *x* under conditions *y*, a certain percentage of subjects give answer *z*. Such data can, at best, operate only at the first of Marr's (1982) three levels, the theory of the computation – that is, an input/output account of what function the system

[12] In addition to its original home in linguistics, the distinction has also done important work in other parts of cognitive science, e.g., in the developmental folk psychology literature (Surian and Leslie 1999; Bloom and German 2000; Scholl and Leslie 2001).

computes. However, explanations in terms of performance error most plausibly operate at either the second or third of Marr's three levels of explanation – the level of the algorithm, and the level of the physical implementation. Without operating at those levels, the appeal to performance errors can be used, at best, to explain away small and unsystematic variation. But the negative experimental philosophy findings present a challenge precisely in virtue of their systematic nature.

To deploy the distinction as a response to those findings would, thus, depend on a construal of the actual workings of the system in question.[13] One cannot separate competence from performance with only input/output data; rather, one requires, at least in the background, some sort of account as to what the idealized operation of the system is supposed to be like, such that performance errors can be explained away in terms of the system falling short of that idealization in some way. In the absence of any processing or physical accounts, we just cannot know how the requisite idealization is supposed to go. Such explanations can only succeed, though, given a reasonably clear idea of what resources are being strained, and preferably also how that resource might be limited in the first place. This is why one standard performance error account of center embeddings works so well – working memory has a pretty good track record in both regards, as revealed in the general popularity of "cognitive load" as an experimental manipulation. But these are not questions at the level of which inputs produce which outputs, for they require some story about the inner workings of the system. They are therefore not the sorts of questions that can be addressed via survey methods.

It may be easier for positive experimental philosophy to apply a competence/performance distinction in terms of one process interfering with another, and this is indeed what we see with Nichols and Knobe (2007).[14] Nichols and Knobe explored two different factors that could influence subjects' willingness to attribute the possibility of moral responsibility in a hypothetical case: (i) whether or not the case was described as being in a deterministic universe or an indeterministic one, and (ii) whether or not the case was affectively engaging.[15] Unsurprisingly, Nichols and Knobe found that subjects were generally more willing to attribute the possibility of responsibility in indeterministic universes than in deterministic ones. Perhaps more unexpectedly, they found a similar increased willingness in high-

[13] It also presupposes at a minimum that it will be possible to decompose the relevant cognition into mechanisms with individually discernible functions. We note this commitment without taking issue with it here.

[14] To our knowledge, no one has explored "physical breakdown" as a candidate source of performance errors in positive experimental philosophy.

[15] We grant here, for the sake of discussion, that they have correctly characterized the way their experimental materials map into these distinctions.

affect cases over low-affect ones. So in a low-affect, deterministic case only a minority of subjects (23%) judged moral responsibility to be possible. A slender majority of subjects judged moral responsibility possible in the high-affect, deterministic case (64%), with more robust majorities in the low-affect, indeterministic case (89%) and most of all in the high-affect, indeterministic case (95%).

So we have some diversity of intuitions across different cases here. While philosophers may not worry about whether or not the determinism/indeterminism differences track a real difference in subjects' competence in attributing agency, they may worry about what to make of the influence of affect on these judgments. Nichols and Knobe consider two possible interpretations. On the "affective competence" interpretation, our emotions are properly part of our agency-attribution system, and their tendency towards compatibilism (as revealed by the majority response in the high-affect, deterministic case), thus, reveals a real commitment of our psychology of responsibility. On the "affective performance error" interpretation, our emotions interfere with the more properly incompatibilist judgments of agency. Here's how they argue for the latter interpretation:

> We think that the affective performance error model provides quite a plausible explanation of our results. What we see in the [low affect] case is that, when affect is minimized, people give dramatically different answers depending on whether the agent is in a determinist or indeterminist universe. On the performance error hypothesis, these responses reveal the genuine competence with responsibility attribution, for in the low affect cases, the affective bias is minimized. When high affect is introduced… the normal competence with responsibility attribution is skewed by the emotions; that explains why there is such a large difference between the high and low affect cases in the determinist conditions.
>
> Now let's turn to the affective competence account. It's much less clear that the affective competence theorist has a good explanation of the results. In particular it seems difficult to see how the affective competence account can explain why responses to the low-affect case drop precipitously in the determinist condition, since this doesn't hold for the high affect case. Perhaps the affective competence theorist could say that low affect cases … fail to trigger our competence with responsibility attribution, and so we should not treat those responses as reflecting our normal competence. But obviously it would take significant work to show that such everyday cases of apparent responsibility attribution don't really count as cases in which we exercise our competence at responsibility attribution. Thus, at first glance, the performance error account provides a better explanation of these results than the affective competence account. (p. 676)

Yet there is a problem here. To describe one process as interfering with another presupposes an individuation of the processes involved, which is again not something that can be done purely with the sort of survey data that Nichols and Knobe have (like almost all practitioners of experimental

philosophy). If we already possessed a well-worked out account of the particular mechanisms operating in these domains and their various interactions, then such an account could maybe provide a framework within which such studies could do the required work. But no such account is currently on offer that can help tell us whether the affective influence on people's judgments a component part of, or an extraneous to, the system producing those judgments. In the absence of any individuation of mechanisms at either the algorithmic or the implementational level of explanation, we cannot tell. The question becomes particularly messy for Nichols and Knobe, as they want to opt for a "hybrid" account in which some of the affect is *part* of the competence, while other parts of it present an interfering factor. We are not arguing about the truth or falsity of *that* claim; but we are addressing whether positive experimental philosophy's survey methods are sufficient to establish such claims as true or false, and we are concerned that they cannot.[16]

One way to see this problem more sharply is to notice that the two hypotheses that Nichols and Knobe consider do not compete only with each other, but must also compete with a number of other hypotheses in which some set of the observed performance is produced by one system and some other set by a different, interfering system. They offer one way of carving things up, but the worry that we're articulating here is that it is extremely difficult, given only the sort of data that they have, to prefer any one of those ways over other possibilities. For example, 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; perhaps the description of the determinist universe triggers some sort of explanation-detection system, which competes with the responsibility-attribution system, and produces improper interference. Or perhaps there is just one unified mechanism, and the profile of responses they report is simply the result of its computations, and there are no performance errors to be explained away at all! Such a result would be *philosophically* surprising, but there is no *psychological* reason, given only survey data, to rule it out.

D The Proper Domain Problem

One way to get around the worries just articulated would be to already possess a mature theory of the computation for the system in question. Although such a theory tells us what the input-output function is for the sys-

[16] See Scholl (2007) for a positive example of using implicit measures in experiments in the philosophical domain of the metaphysics of objects.

tem, it does more than that – it tells us what function it is that we should understand the system as computing. If we already had such a theory, then we could use it to help referee between at least some competing accounts at the algorithmic level. This could perhaps be put to use to separate conceptual from non-conceptual components of cognition, or to help underwrite an application of the competence/performance distinction to discern the boundaries of a computational mechanism.

Although such an appeal to the theory of the computation is theoretically possible, it will not help here. For we are considering cases in which positive experimental philosophy is supposed to help *referee* between conflicting accounts of a given domain. As such, these are cases in which the fundamental philosophical facts are dialectically up for grabs. To determine whether Nichols and Knobe's subjects' affect is interfering with their judgments, or is, instead, a manifestation of their competence at making those judgments, we need to know first what function it is that their psychological systems are trying to compute. If their systems are meant to judge the world along the lines of compatibilism, then the affect would be part of the competence; and if their systems are meant to judge the world along the lines of incompatibilism, then the affect would be an interference.

(Note that this worry, perhaps unlike the worry in the previous section, applies more generally than just to survey-based positive experimental philosophy. For example, it applies equally well to Joshua Greene's attempts to argue for utilitarianism based on his neuroimaging studies of subjects considering trolley cases (e.g., Greene *et al*. 2001). We fear that practitioners of positive experimental philosophy have overlooked the extent to which successful psychological appeals to a competence/performance distinction typically rely on key theoretical factors such as (i) substantial theories of the system in question, (ii) an account of the resources the system will need to draw upon, and (iii) a specification of the proper domain of the system. And when such theories are not in fact available, then appeals to that distinction are easily challenged (see, e.g., Stanovich and West 2000, section 3). We unfortunately lack such theories in most areas of positive experimental philosophy.)

Practitioners of positive experimental philosophy might, thus, look around for other sources of evidence that could help determine what function it is that, say, our "moral responsibility" system is computing. But this will turn out to be difficult to do: there is no reason to think, even conceding that the mind is comprised of systems discrete enough to be assigned separate proper domains, that such systems will correspond closely enough with domains of philosophical interest (e.g. moral responsibility) to provide acceptable reductions of them. And to lose track of this fact is to distort both the philosophy and the psychology.

Philosophers are typically interested in psychological domains that cor-

respond to philosophical concepts of interest, for example, *responsibility, morality, knowledge,* and so forth. But there is no reason to think that these concepts neatly align with, rather than cross-cut, our cognitive architecture. To see this, consider recent claims of a "linguistic analogy" for the moral domain – claims that moral cognition is underwritten by a domain-specific adaptation for morality, on a par with the linguistic faculty posited by Chomsky (e.g., Dwyer 1999; Harman 1999; Mikhail 2000; and Hauser *et al.* 2007).

One way of thinking about proper functions is in terms of evolution: the proper function of a mechanism is the function for which the mechanism was, or is, selected for – the function that computes solutions to problems in its *proper domain* (Sperber 1994). While there could be other ways of specifying what exactly determines or constitutes the proper function of a mechanism, the main point is that such proper functions may either be directed at problem domains that are more or less neatly coextensive with philosophical domain of interest like morality or they may not. Indeed, some proponents of a moral faculty are quite clear about this (see, e.g., Hauser *et al.* 2007): they mean to make the nontrivial claim that there is a species-typical, innate, and domain-specific mechanism whose proper function is moral judgment, and they understand such proper functions as produced by evolutionary pressures to solve moral problems.

But when we start thinking about natural selection, it suggests pressures that lead in quite different directions than moral reasoning. To choose a quick and straightforward example, it seems like sound evolutionary logic to think that evolution might favor mechanisms that systematically favor members of one's own family, or members of one's own group, or persons that might assist one's reproductive success, while it is at least plausible to think that these considerations are not *morally* relevant. At the very least, one cannot simply assume that the domain determined by the proper function of the mechanism or mechanisms that underwrite morality and the domain of morality are coextensive.[17]

Consider how losing track of the distinction between the domain of morality and the proper domain of the mechanism that implements morality can distort psychological inquiry. Much recent work in moral psychology, including work in the "linguistic analogy" tradition, has engaged in relatively straightforward appropriation of the philosophical technique of eliciting moral intuitions by presenting carefully constructed moral dilemmas (e.g., Mikhail 2000; Hauser *et al.* 2007). This appropriation, however, makes the substantial assumption that the kind of data of relevance to philosophers will also be relevant to psychologists. Notice, first, that philoso-

[17] For further elaboration on this point, see Mallon (2007).

phers carefully construct moral dilemmas as ways of eliciting intuitions of relevance to assessing competing theories of morality. For this reason, philosophers' dilemmas typically *exclude* factors that are widely considered to be morally irrelevant. For example, variations of the famous "runaway trolley" typically aim to probe the circumstances (if any) in which it is okay to bring about the death of one person in order to save five others. But such variations typically do not include versions in which we are asked to weigh the lives of relatives, out-group members, potential sex partners, and so forth, and (to repeat), this is precisely because whatever difference such factors might make is irrelevant to the moral questions at hand. But, there's every reason to think that these factors *are evolutionarily* relevant and so considering them in computing moral judgments may be part of the proper function of whatever mechanism or mechanisms that underlie moral judgment.

When we deviate from this assumption and consider different factors, we may find different answers – answers that suggest a very different sort of faculty than a "moral" faculty is at work. In one of the earliest experimental investigations of trolley dilemmas, Petrinovich, O'Neill, and Jorgensen (1993) report finding that subjects prefer the lives of relatives and friends over strangers in standard trolley scenarios, a finding they take to support sociobiologists' and evolutionary psychologists' suggestions that humans are designed, in part, to be concerned with their own inclusive fitness. Suppose that this data from Petrivonich *et al.* is correct. Suppose further that much of our moral judgment is underwritten by an evolutionarily designed mechanism *M* that computes using an internalized principle like:

> (K) The wrongness of an action resulting in an avoidable death is inversely proportional to the subject's relatedness to me.

Such a principle may well be *morally* irrelevant, but at the same time it may well be central to the operation of the faculty that underlies moral judgments about trolley cases. Hauser *et al.* (2007) indicate that, in contrast with such research that focuses on questions of "evolutionary significance," their research will probe "the computational operations that drive our judgments" (p. 127). But this begs a crucial question, viz. whether the computational process driving our typical moral judgments are themselves biased by evolution in ways that are at odds with our intuitive sense of morality.[18] It

[18] Hauser *et al.* make this same move more explicitly when they exclude gender as a relevant explanatory dimension, writing that "we find it clear that some distinctions (e.g., the agent's gender) do not carry any explanatory weight" (Hauser *et al.* 2007, p. 131). Here again, they make judgments that reflect a judgment about what sort of considerations are properly considered moral ones. But there seems little reason to think evolution would have respected such niceties in constructing us, so it is not clear why such exclusions are relevant to our underlying functional organization.

is possible that our concept of morality may emerge only when an innate, domain specific mechanism is used in ways that are at odds with its design (e.g. when it is not allowed access to information such as the relatedness of a person to us).

If philosophical and psychological boundaries needn't be even approximately isomorphic, then it goes to show that inquiries into those borders are relatively autonomous. One cannot read the borders of philosophically interesting domains off of the psychology, and one cannot read psychological borders off of the philosophy.

4 Conclusion & Prospects for the Future

Let us recap how the worries we have raised here can be seen to afflict the particular positive programs. Since all of the positive programs are committed to the view that intuitions are trustworthy and shared, each faces the challenge of figuring out what to do with the results of negative experimental philosophy. In particular, the positive programs owe us the same kind of story that traditional armchair philosophers owe: how do we discern which intuitions count? When different folks yield different intuitions, which one do we take to be likely to be tracking the philosophical truth?

This challenge can make some sort of mentalism attractive. One resource that mentalism offers is the legitimacy of some degree of relativism: maybe people with different intuitions just have different concepts, so everyone is still correct. (Here we see a clear case where positive experimental philosophy has an advantage over traditional armchair methods: the latter has no capacity to discern such demographic differences in intuitions, and hence concepts, whereas the former can commission as much cross-cultural research as is needed.) But, we are skeptical about just how attractive, or how helpful, such a move towards relativism is. Another resource that mentalism offers is some means for explaining *away* some of the variation, as due to factors other than the meaning-constitutive elements of the concepts themselves. But appealing to this resource entangles mentalism in Quinean difficulties that it has not (and, we worry, cannot) resolve. Finally, mechanist mentalism offers the possibility of accommodating the results of negative experimental philosophy by accounting for the various forms of inter- and intra-personal instability and variation in terms of performance errors. Mechanist mentalism, however, seems to require *either* a well worked-out architecture of the different systems involved, *or* some way of fixing the proper domain of those systems. Given the kinds of survey data typically gathered in

current positive experimental philosophy research, though, and the problem of philosophical domains cross-cutting psychological domains, mechanist approaches seem to lack both the experimental and theoretical tools needed to advance their programs at this time.

At this point, the prognosis might seem rather grim for experimental philosophy's positive programs. Interestingly, their salvation might ultimately rest on their ability to become *more* experimental – or at least more like experimental psychology. The kinds of survey methods that experimental philosophers so frequently employ play little role in experimental psychology – and, for good reason: there are better methods available to answer the kinds of questions that are of interest to both experimental psychologists and philosophers (Scholl 2007).[19] Experimental philosophy's positive programs would do well, we think, to become more like experimental psychology. The hope of experimental philosophy's positive program was to use science to help do some of the work that traditional philosophy hasn't been able to do (or hasn't been interested in doing). Part of the challenge facing the positive programs is to become more scientifically sophisticated.

But doing more, and better, *science* will not be enough by itself to fully met the challenge – the positive programs must also do enough *philosophy* to see how to bridge the gap from empirical findings to philosophical payoffs. Positive experimental philosophy may face a situation now that Wittgenstein asserted once of psychology: "there are experimental methods and *conceptual confusion* ... The existence of experimental methods makes us think we have the means of solving the problems which trouble us; though problem and methods pass one another by."[20] If we are right that neither psychologists nor linguists are really in the business of providing answers to the kind of meaning-constitutivity questions that semanticists and conceptualist philosophers are interested in, then philosophers will have to figure out for themselves how empirical results can speak to such questions. And we expect that attempts to draw a competence/performance distinction in such areas as moral or epistemic judgment, as many mechanist philosophers have been interested in doing, cannot succeed

[19] Scholl provides some excellent suggestions as to how philosophers and psychologists could do a better job of getting a handle on the mechanisms underlying various intuitions. However, it is not clear how to use his suggestions to help with the deeper sorts of problems discussed here. For example, he writes, "understanding the origins of our metaphysical intuitions in various psychological mechanisms could help us understand when they are worth revising or forfeiting in our philosophical theories, especially if there is reason to think that those psychological mechanisms may yield unreliable results in the particular contexts in which they are being asked to operate" (Scholl 2007, p. 586). But without knowing the proper domain of those mechanisms, it stands as an open question just what will count as reliable or unreliable operation.

[20] Wittgenstein (1953), Part II, section xiv.

without drawing on at least some substantive philosophical claims about the nature of those domains. Compare, for example, how impossible a theory of 3-D vision would be without drawing on claims about the structure of the physical world about which vision gives us information, and in particular the way that visible light interacts with that world. Finally, we see no reason to think that appeals to the *proper domain* of a mechanism will gain much traction since there's no reason to think that, even granting that cognitive mechanisms are domain-specific, the proper domains of cognitive mechanisms will be isomorphic to domains of philosophical interest. Positive experimental philosophy may, in the end, be able to avoid some of the methodological pitfalls of its armchair predecessors. But it seems that some challenges are endemic to philosophy *qua* philosophy, no matter where one sits while one does it.

References

Adams, F., and Steadman, A. 2004a. Intentional Action in Ordinary Language: Core Concept or Pragmatic Understanding? *Analysis* 64:173-181.

Adams, F., and Steadman, A. 2004b. Intentional Action and Moral Considerations: Still Pragmatic. *Analysis* 64:268-276.

Alexander, J., and Weinberg, J. 2007. Analytic Epistemology and Experimental Philosophy. *Philosophy Compass* 2:56-80.

Bloom, P., and German, T.P. 2000. Two reasons to abandon the false belief task as a test of theory of mind. *Cognition* 77:B25-B31.

Christensen, D. 2007. Epistemology of Disagreement: The Good News. *The Philosophical Review* 116:187-217.

Cummins, R. 1998. Reflection on Reflective Equilibrium. DePaul, M, and Ramsey, W. (eds.), *Rethinking Intuition* (pp. 113-28). Lantham: Rowman and Littlefield Press.

Dwyer, S. 1999. Moral Competence. Murasugi, K., and Stanton, R. (eds.), *Philosophy and Linguistics* (pp. 169-90). Boulder: Westview Press.

Elga, A. 2006. Reflection and Disagreement. *Nous* 41:478-502.

Feldman, R. 2006. Epistemological Puzzles About Disagreement. Heatherington, S. (ed.), *Epistemology Futures*: Oxford: Oxford University Press.

Feldman, R., Warfield, F. 2007. *Disagreement*. Oxford: Oxford University Press.

Gelman, S. 2003. *The essential child: Origins of essentialism in everyday thought*. Oxford: Oxford University Press.

Glasgow, J. 2008. On the Methodology of the Race Debate: Conceptual Analysis and Racial Discourse. *Philosophy and Phenomenological Research* 76:333 – 358.

Goldman A., and Pust, J. 1998. Philosophical Theory and Intuitional Evidence. DePaul, M, and Ramsey, W. (eds.), *Rethinking Intuition* (pp. 179-200). Lantham: Rowman and Littlefield Press.

Greene, J., Sommerville, R., Nystrom, L., Darley, J., and Cohen, J. 2001. An fMRI Investigation of Emotional Engagement in Moral Judgment. *Science* 293: 2105-2108.

Greene, J. 2003. From neural "is" to moral "ought": what are the moral implications of neuroscientific moral psychology? *Nature Reviews Neuroscience* 4:847-850.

Harman, G. 1999. Moral philosophy and linguistics. Brinkmann, K. (ed), *Proceedings of the 20th World Congress of Philosophy, vol. I: Ethics* (pp. 107-15). Bowling Green, OH: Philosophy Documentation Center.

Hauser, M., Young L., and Cushman, F. 2007. Reviving Rawls' Linguistic Analogy: Operative principles and the causal structure of moral actions. Sinnott-Armstrong, W. (ed), *Moral Psychology, Volume 1: The Evolution of Morality: Adaptations and Innateness.* Cambridge, MA: MIT Press (Bradford Books).

Jackson, F. 1998. *From Metaphysics to Ethics: A Defense of Conceptual Analysis*. Oxford: Oxford University Press.

Johnson, K. 2008. An Overview of Lexical Semantics. *Philosophy Compass* 3:119–134.

Kauppinen, A. 2007. The Rise and Fall of Experimental Philosophy. *Philosophical Explorations* 10:95-118.

Kelly, T. 2005. The Epistemic Significance of Disagreement. Hawthorne, J., and Gendler Szabo, T. (eds.), *Oxford Studies in Epistemology, Volume* 1 (pp. 167-96). Oxford: Oxford University Press.

Kelly, T. 2007. Peer Disagreements and Higher Order Evidence. Feldman, R., and Warfield. T. (eds.), *Disagreement*. Oxford: Oxford University Press.

Kelly, T. 2008. Disagreement, Dogmatism, and Belief Polarization. *The Journal of Philosophy* 105:611-633.

Knobe, J. 2003a. Intentional Action and Side Effects in Ordinary Language. *Analysis* 63:190-193.

Knobe, J. 2003b. Intentional Action in Folk Psychology: An Experimental Investigation. *Philosophical Psychology* 16:309-324.

Knobe, J. 2004. Intention, Intentional Action and Moral Considerations. *Analysis* 64:181-187.

Knobe, J. 2007a. Experimental Philosophy and Philosophical Significance. *Philosophical Explorations* 10:119-122.

Knobe, J. 2007b. Reason Explanation in Folk Psychology. *Midwest Studies in Philosophy* 31:90-107.

Lewis, D. 1970. How to define theoretical terms. *Journal of Philosophy* 67:426-446.

Lewis, D. 1972. Psychophysical and theoretical identifications. *Australasian Journal of Philosophy* 50:249-58.

Machery, E. 2008. The folk concept of intentional action: Philosophical and psychological issues. *Mind & Language* 23:165-189.

Machery, E. 2009. *Doing without concepts*. New York: Oxford University Press.

Machery, E., Lindquist, S., and Griffiths, P. Forthcoming. The Vernacular Concept of Innateness. *Mind & Language*.

Machery, E., Mallon, R., Nichols, S., and Stich, S. 2004. Semantics, Cross-cultural style. *Cognition* 92:B1-B12.

Mallon, R., Machery, E., Nichols, S., Stich, S. Forthcoming. Against Arguments from Reference. *Philosophy and Phenomenological Research.*

Marcus, M. 1980. *Theory of syntactic recognition for natural languages*. Cambridge: MIT Press.

Mikhail, J. 2000. Rawls' linguistic analogy: A study of the 'generative grammar' model of moral theory described by John Rawls in 'A theory of justice'. Ph.D. Thesis. Cornell University, Ithaca.

Mallon, R. 2007. Reviving Rawls Inside and Out. Sinnott-Armstrong, W. (ed), *Moral Psychology, Volume 2: The Cognitive Science of Morality: Intuition and Diversity* (pp. 145-55). Cambridge MA: MIT Press (Bradford Books).

Mallon, R., and Nichols, S. Forthcoming. Moral Reasoning, Moral Rules, and Moral Dilemmas. Doris, J., Nichols, S., and Stich, S. (eds.), *The Oxford Handbook of Moral Psychology*. Oxford: Oxford University Press.

Nadelhoffer, T. 2004a. On Praise, Side Effects, and Folk Ascriptions of Intentionality. *Journal of Theoretical and Philosophical Psychology* 24:196-213.

Nadelhoffer, T. 2004b. Blame, Badness, and Intentional Action: A Reply to Knobe and Mendlow. *Journal of Theoretical and Philosophical Psychology* 24:259-269.

Nadelhoffer, T., and Nahmias, E. 2007. The Past and Future of Experimental Philosophy. *Philosophical Explorations* 10:123-149.

Nahmias, E., Morris, S. Nadelhoffer, T., and Turner, J. 2005. Surveying Freedom: Folk Intuitions about Free Will and Moral Responsibility. *Philosophical Psychology* 18:561-584.

Nahmias, E., Morris, S., Nadelhoffer, T., and Turner, J. 2006. Is Incompatibilism Intuitive? *Philosophy and Phenomenological Research* 73:28-53.

Nichols, S. 2002. On the Genealogy of Norms: A Case for the Role of Emotion in Cultural Evolution. *Philosophy of Science* 69:234-255.

Nichols, S. 2006. Imaginative Blocks and Impossibility: An Essay in Modal Psychology. Nichols, S. (ed.), *The Architecture of the Imagination* (pp. 237-55). Oxford: Oxford University Press.

Nichols, S., and Knobe, J. 2007. Moral Responsibility and Determinism: The Cognitive Science of Folk Intuitions. *Nous* 41:663-685.

Nichols, S., and Mallon, R. 2006. Moral Rules and Moral Dilemmas. *Cognition* 100:530-542.

Nichols, S., Stich, S., and Weinberg, J. 2003. Metaskepticism: Meditations in Ethno-Epistemology. Luper, S., (ed.), *The Skeptics: Contemporary Debates* (pp. 227-47). Burlington, VT: Ashgate Press.

Nichols, S., and Ulatowski, J. 2007. Intuitions and Individual Differences: The Knobe Effect Revisited. *Mind and Language* 22:346-365.

Petrinovich, L., O'Neill, P., and Jorgensen, M. 1993. An Empirical Study of Moral Intuitions: Toward an Evolutionary Ethics. *Journal of Personality and Social*

Research 64:467-478.

Petrinovich, L., and O'Neill, P. 1996. Influence of wording and framing effects on moral intuitions. *Ethology and Sociobiology* 17:145-171.

Pizzaro, D., Uhlman, E., Tannenbaum, D., and Ditto, P. Manuscript. The Motivated Use of Moral Principles.

Quine, W. 1951. Two Dogmas of Empiricism. *Philosophical Review* 60:20-43.

Scholl, B. 2007. Object Persistence in Philosophy and Psychology. *Mind & Language* 22:563-591.

Scholl, B., and Leslie, A. 2003. Minds, Modules, and Meta-Analysis. *Child Development* 72:696-701.

Schwitzgebel, E. Forthcoming. Do Ethicists Steal More Books. *Philosophical Psychology*.

Sperber, D. 1994. The Modularity of Thought and the Epidemiology of Representation. Hirchfeld, L., and Gelman, S. (eds.), *Mapping the Mind* (pp. 39-67). Cambridge: Cambridge University Press.

Stotz, K., and Griffiths, P. 2004. Genes: Philosophical analyses put to the test. *History and Philosophy of the Life Sciences* 26:5-28.

Surian, L., and Leslie, A. 1999. Competence and performance in false belief understanding: A comparison of autistic and normal 3-year-old children. *British Journal of Developmental Psychology* 17:141-155.

Swain, S., Alexander, J. and Weinberg, J. 2008. The Instability of Philosophical Intuitions: Running Hot and Cold on Truetemp. *Philosophy and Phenomenological Research* 76:138-155.

Ulatowski, J. 2008. How Many Theories of Act Individuation Are There? Ph.D. Thesis. University of Utah, Salt Lake City.

Weinberg, J. 2007. How to Challenge Intuitions Empirically Without Risking Skepticism. *Midwest Studies in Philosophy* 31:318-343

Weinberg, J., Nichols, S., and Stich, S. 2001. Normativity and Epistemic Intuitions. *Philosophical Topics* 29:429-60.

White, R. 2005. Epistemic Permissiveness. *Philosophical Perspectives* 19:445-459.

Williamson, T. 2004. Philosophical 'Intuitions' and Scepticism about Judgments. *Dialectica* 58:109-155.

Williamson, T. 2005. Armchair Philosophy, Metaphysical Modality and Counterfactual Thinking. *Proceedings of the Aristotelian Society* 105:1-23.

Williamson, T. 2007. *The Philosophy of Philosophy*. Oxford: Blackwell Publishing.

Wittgenstein, L. 1953. *Philosophical Investigations*. Oxford: Blackwell Publishing.