
Two Kinds of Experimental Philosophy (and their methodological dangers)

Brian Scholl, Yale University
brian.scholl@yale.edu
<http://www.yale.edu/perception/>

These are some brief, informal, and probably impertinent notes based on a talk given on 6/25/08 at the Pre-SPP Workshop on Experimental Philosophy. I received some requests for the Keynote slides I used during this talk, but those would really not be helpful to others, as they were just a haphazard mix of visual aids. I have no plans to write a paper based on that material, though, so as a compromise I jotted down this quick outline of some of the thoughts that were covered in that talk.

1. Overview

What is experimental philosophy (X-phi)? Here's one answer: "Experimental philosophers proceed by conducting experimental investigations of the psychological processes underlying people's intuitions about central philosophical issues" (Knobe & Nichols, 2008). This isn't the only way to characterize X-phi, but from my perspective it's the most important and interesting one. Interacting with philosophy as an outsider of sorts, it seems clear that a great deal of it relies on *intuitions* (often about particular situations and puzzles), but that there has traditionally been an odd lack of attention to where those intuitions come from. X-phi seems to me to be the most successful and prominent attempt to fill this gap. And I should be clear that I think that much of this work is fantastic — indeed, among the most interesting work being done at the intersection of philosophy and psychology today.

Nevertheless, I think X-phi as it is typically conducted almost never succeeds in practice — and is unlikely to succeed in principle — at the goal stated above: determining the nature of the processes that produce the relevant intuitions. The primary reason for this failure is that the methods that are almost-but-not-quite-universally used in this area — viz. "survey methods" — are spectacularly ill-suited to this goal.

2. The Methods of X-phi

Traditional analytic philosophy makes use of "thought experiments". X-phi is meant to contrast with this, but in my view the dominant methods of X-phi seem to be ... thought experiments. The only difference is that in X-phi the thoughts are aggregated, and said out loud, or written down. This isn't how these methods are usually characterized, but that seems to be their essence. They are sometimes called "survey methods" (I ask you and your friends about your thoughts), "vignette studies" (I make you read about some weird things before asking you about your thoughts), or "rating studies" (I force you to convert your thoughts into a number along a specified continuum) — but in the end these are really just dressed-up thought experiments.

3. 4 Challenges

I worry that these sorts of methods end up telling you less about the psychological processes that underlie the philosophically relevant intuitions, and more about the psychology of how people cope with being asked strange questions. The reason for this is that many different and often irrelevant cognitive processes can readily contribute to the answers that subjects give to these questions. (As will be noted below, this is less true of other methods.) This general problem has been noted in many different contexts in actual X-phi debates. Let's review:

Problem #1: Subjects often fail to interpret the questions in the way that you intend

This has been a salient type of critique in practice for many X-phi studies. Take, for example, the "Knobe Effect"; most of the published critiques of this fascinating phenomenon have really ended up

complaining that subjects might be answering the wrong question. When a subject says “Sure, the chairman *intentionally* polluted the environment” via a survey answer or a free response, we can imagine the underlying reasoning that we *hope* underlies this answer: “Hmm. Now that I think about, it really does seem that he did that intentionally. So yes, I’ll choose that answer.” But we can easily imagine others, such as: “That chairman: what a bastard. Boy, I wish I could stick it to him. What would be the best way to do that, I wonder? Damn, there’s no ‘punish him!’ option. Hmm; which of these other weird options would be most likely to lead to punishment. ‘Intentionally’, I suppose; I’ll choose that.” (cf. Adams & Steadman, 2004). How common is this? Well, we don’t know, since most X-phi “experiments” simply collect the responses themselves. Very occasionally, experimenters do ask subjects in more detail about what went into their answers, and the results of such followup questions should be a cause for great concern. E.g.: “Of the 30 participants, 13 gave explanations that indicated that they failed to abide by the conditions of the thought experiment” (Nichols, 2006, *Journal of Cognition & Culture*). (Just imagine if this was true for all of the X-phi studies that *didn’t* collect post-hoc explanations!)

Problem #2: Even when Ss do interpret the questions in the way you intend, their answers will depend on a host of extrinsic contextual factors

Responses to survey questions can be (and are almost always) influenced by many different extrinsic factors beyond the actual intuitions themselves. There are too many to mention, but they include order effects (as nicely demonstrated by Jonathan Weinberg’s group), arousal, inferences about authority, priming, and many others. Worse, the influence of these specific contextual factors can differ by culture, SES, etc. (meaning that diversity in responses to such questions — à la the ‘negative program’ — may not reflect the underlying intuitions themselves either). I suspect that that even the knowledge that they are in a ‘philosophy’ study could make a difference here (cf. Norenzayan & Schwarz’s infamous 1999 “letterhead” studies, wherein they obtained reliable differences in survey responses based only on whether the letterhead at the top of the survey said *Institute for Social Research* or *Institute for Personality Research*). All of these factors are irrelevant to the underlying intuitions that the experimenters are hoping to capture, but survey methods do a very poor job (in contrast to other performance-based methods) of screening them out.¹

Problem #3: Worse yet, often the (forced-choice) methods will yield an answer even the subject has no intuition at all — and it’s extremely difficult to design such studies w/o some confounding ‘surface’ variable

Suppose that as a subject in one of these survey experiments you simply didn’t have a strong intuition about the question at all. Often task demands would still implicitly force you to give an answer — and often that answer would tend in a reliable direction because of meta-experimental factors. Here are two examples.

The first is from some psychologists attempting to get at philosophical intuitions about object persistence (Rips, Blok & Newman, 2006). They showed subjects Y-shaped figures, described as rivers, wherein one ‘fork’ of the Y was more similar to its ‘stem’ (e.g. being drawn with the same stroke-width), while the other fork was less featurally similar (e.g. being drawn with a different stroke-width). Subjects were asked about which fork was more likely to be the continuation of the same river. Subjects were more likely to choose the fork that was more similar to the stem, and the authors conclude that their studies “provide evidence about which of the experimental factors affect decisions about individual identity”. I doubt it. Suppose you had no intuition about this type of “which is the same river” question (as I do); how would you respond? To always give the same answer to every probe would seem like you’re being an uncooperative scoundrel. To answer each probe randomly would also seem like you’re being an uncooperative scoundrel. So you have to answer systematically in some way. But how to do that? Well, there is a single salient feature that is obviously being varied across the questions (within-subjects), so it seems natural to vary your responses according to that variable. And even if you see no link between similarity and persistence,

¹ I’m assuming here that there *are* some ‘pure’ intuitions that aren’t contaminated by such extrinsic factors, but Ron Mallon’s excellent talk at this same workshop made me doubt even that!

everyone is going to have the same meta-experimental intuition about which direction the effect *would* go if there was such an effect.

Another great example comes from Knobe & Prinz's (2008) X-phi study of whether group entities can have phenomenal states. Their subjects were presented with statements such as "*Acme Corp. wants to change its corporate image*" (which attributes a non-phenomenal mental state to the corporation) vs. "*Acme Corp. is experiencing a sudden urge to pursue internet advertising*" (which attributes a phenomenal state to the corporation) — and they simply had to rate these statements on a scale from 1 ("sounds weird") to 7 ("sounds natural"). You already know the results of this study, since it works as a thought experiment too: it sounds weirder to attribute phenomenal states to group entities. But this doesn't seem like a compelling experiment, since you'd get the same patterns of results even if you didn't have any such intuitions. Again, in order to avoid being a scoundrel, you can't simply answer "3" for every statement, and you can't simply answer randomly. But wait: again there's an obvious variable being manipulated within-subjects (whether "feeling" words are used), so it can be natural to vary your responses according to that variable (and in the obvious way), even if you didn't have such actual intuitions. (And note, by the way, that simply asking subjects to explain the reasons behind their responses — cf. Nichols & Ulatowski, 2007 — won't necessarily highlight this problem unless the follow-up questions are asked very carefully, since subjects may simply advert to the superficial reasons that guided their responses when asked "Why did you give those answers?").²

Problem #4: Even when these sorts of methods succeed at drawing out intuitions, they're rarely able to determine where they came from

X-phi is meant to contrast with "armchair methods" in large part because those methods have limitations: "From the armchair we can't discern which psychological mechanisms subserve the intuitions. Nor can we determine whether conflicting intuitions have their origins in different psychological mechanisms" (Nichols, 2006). But, in my view, the 'survey methods' of X-phi have these same limitations: even if they succeed at drawing out intuitions, they almost never succeed in practice at isolating the responsible cognitive processes. These methods simply don't have that grain of resolution. (I note that in Nadelhoffer & Nahmias' 2007 review, they highlight this as one of the key goals of X-phi, but then they don't cite a single example of survey methods that have met this goal — instead focusing on Josh Greene's initial fMRI work, etc.)

4. Interim Summary

Of course, it is possible to combat these types of problems, and X-phi-ers are doing so: it has become more commonplace to ask careful followup questions, e.g., and within-subjects studies are becoming less common. There is also a whole literature on how to do these types of studies well, including several edited books (e.g. Krosnick & Fabrigar, in press, *Handbook of Questionnaire Design*; Sirken et al., 1999, *Cognition and Survey Research*) and special issues of journals such as *Applied Cognitive Psychology* (e.g. the 1991 special issue on *Cognitive aspects of survey methodology* and the 2007 special issue on *Cognitive psychology & survey methods: Nurturing the dialogue...*). Even so, I note that (1) such challenges are *endemic* to these types of methods, and (2) these are the types of challenges that in practice have fueled many of the internal disputes in X-phi about particular results. The methods can be improved, but only to make such studies progressively "less bad". (I'm being intentionally provocative here, of course, but I'm also trying to express a genuine deep concern with this sort of "survey" study.) These types of methods — aggregating thought experiments and either saying the results out loud or writing them down — simply don't contrast *enough* with armchair philosophy, and they are highly imperfect ways even in principle to probe the psychological processes that underlie philosophical intuitions.

² Let me stress that I really like the phenomena that Knobe & Prinz have discovered and highlighted in this study, and I believe it. But I believe it not because it's a compelling experiment, but because it's a compelling *thought experiment!* I have this reaction to many of X-phi's greatest hits. In fact, here's a blasphemous thought: I suspect that the "Knobe effect" would also have been just as prominent and important had it not been introduced in the context of X-phi. What makes that case so interesting is not anything about the experimental results themselves, but just that it's an excellent case, and that the resulting intuition is robust.

5. Another Path

In Nadelhoffer & Nahmias' 2007 review, they note: "Experimental philosophy is the name for a recent movement whose participants use *the methods of experimental psychology* to probe the way people make judgments that bear on debates in philosophy" (my emphasis). As an experimental psychologist, it sounds odd to me to call survey methods "the methods of experimental psychology". Such methods are used sometimes, but rarely. I haven't bothered to check, but my hunch is that if you picked up a random recent issue of each of the top experimental psychology journals, you'd find that < 1% of the papers used any such methods. (They're used more often in social psychology, though — and even more often in other fields such as marketing.) There are good reasons for this, sometimes: there are other performance-based methods that are much more successful in principle and in practice at squeezing answers out of particular underlying cognitive processes.

I think that this would be a useful alternative strategy for X-phi: *Start with the underlying processes in the first place!* Step 1: Catalog the key intuitions and the impact of the key manipulations. Step 2: Test those same questions and manipulations in *focused experimental probes of known underlying processes*. Step 3: Evaluate the matches between these answers: if some underlying processes (but not others) generate the same sorts of answers as do the subjects in your thought experiments, then this lends support to the idea that those are the responsible underlying processes. Some X-phi-ers are starting to do this — e.g. Josh Greene's response-time studies of trolley problems using manipulations of cognitive load, etc.

We have recently tried to implement this strategy in practice. Our case study was *object persistence*: what makes an object the same persisting individual over time? This is obviously a very prominent question in metaphysics. But there is also a vibrant area of psychological investigation, wherein we are able to ask what causes various parts of the mind — especially automatic mechanisms of mid-level visual cognition — to treat an object as the same persisting individual from some previous encounter. It turns out that there are automatic principles of object persistence that are wired into the mind, and they are *primitive* in at least three senses: they operate early in development, they are evolutionarily ancient, and they operate early in online visual processing, before higher-level cognition comes into play. Moreover, the operation of such mechanisms of object persistence turns out to be extremely important for many other cognitive processes, including visual memory, implicit learning, numerical cognition, motion perception, and even visual awareness itself. In our work, we first identified the three most salient themes of philosophical accounts of object persistence in metaphysics, and then we evaluated how these same themes were treated in focused probes of mid-level visual processes. This work looks nothing like survey research: instead, it involves performance-based methods of visual processing such as multiple-object tracking, change detection, object reviewing, etc. — as well as functional neuroimaging. Nevertheless, this study revealed that mid-level visual processing seems to operate according the same general principles that are also the prime movers in philosophical work on persistence in metaphysics. This led us to suggest that some of our key intuitions about persistence that anchor metaphysical theories may in fact derive from mechanisms of mid-level object-based visual cognition. In other words, our intuitions about the nature of persisting individuals may derive from how we *experience* the world in terms of persisting individuals, along with the imagery, simulation, and tacit reasoning that accompany consideration of the relevant thought experiments. In still other words, our judgment about whether an object has persisted as the same individual across some sort of interruption may derive in part from whether our *actual mental representation of that object* (its 'object file') will in fact persist across such interruptions. This very different form of experimental philosophy can yield philosophical progress. For example, if we can learn about where these types of intuitions come from, then we can make more informed choices about which intuitions to give up when a conflict arises.

For a much more detailed summary of this 'other' kind of experimental philosophy, see:

Scholl, B. J. (2007). Object persistence in philosophy and psychology. *Mind & Language*, 22(5), 563 - 591.