Problems and Possibilities for Empirically Informed Philosophy of Mind

The use of empirical work in philosophy of mind is increasing trend, now segueing into philosophy of cognitive science, and the starting point for this chapter. While in favour of this kind of interdisciplinary research, several problems are outlined that raise important questions about the nature of interdisciplinary research across philosophy and the mind/brain sciences. These include how empirical work can be used to support or revise existing philosophical positions, and the role of the empirically-based philosopher in cognitive science. Following this, I suggest an alternative way of approaching questions in philosophy of mind and cognitive science in an interdisciplinary way, based on contemporary work in philosophy of science. This approach is explored through two examples, focusing on the interpretation of first-person data, and questions about the boundaries of cognition. While not the only, or necessarily the best, approach to interdisciplinary work, I suggest that a focus on methodological questions from the point of view of philosophy of science is a potentially invaluable way of pursuing philosophical questions about the mind.

1. Introduction

This chapter is based on the sociological observation that many philosophers interested in the mind now think that interdisciplinary work across philosophy and the cognitive sciences is a Good Thing. While far from uncontroversial in philosophy of mind (e.g. Burge 2010, 2011; McDowell 2010), increasing numbers of philosophers are now interested in psychology and neuroscience (and vice versa), and I take it that this trend will continue. However, I think there are a number of problems facing any attempt to do ‘good’ empirically informed philosophy of mind.

I introduce two of these problems below, that vary from empirical examples simply failing to fit philosophical distinctions (even though they may be used in support of
them), to philosophical accounts taking the form of inadequate scientific theories, in the sense that they are vague and only attempt to fit, rather than predict, empirical data. These problems are certainly not new ones, and have been commented on before with regard to specific cases (some examples are discussed below). Further, the problems outlined here are certainly not intended to support the idea of philosophy of mind as an intellectual enterprise that can, or should, be pursued independently of empirical considerations, or that all philosophy of cognitive science is deeply flawed. Instead, these problems are pointed out as they highlight important methodological questions about the nature of interdisciplinary work. Given the assumption that such interdisciplinary work is likely to continue in philosophy, these problems deserve more discussion.

While these problems are no unresolvable, the second part of this chapter promotes an alternative and potentially very powerful approach to pursuing interdisciplinary philosophical work. This approach is based within philosophy of science: to pursue the philosophical analysis of scientific methods, concepts and distinctions used to investigate the brain. To some extent this is part of the project of philosophy of cognitive science, but I suggest that more general frameworks from philosophy of science (i.e. not those just limited to cognitive science) are so far seriously underused, yet extremely valuable. Two examples of questions that may best be tackled using a philosophy of science approach are discussed, including the analysis of first person or introspective methods, and the sort of constitutive questions that philosophers often seem to assume have a simple answer (e.g. whether cognition is extended or not).

A philosophy of science approach is not the only viable philosophical approach to understanding the mind, nor necessarily the best one, but does offer a way of informing and calming (if not resolving) some existing philosophical debates, and delineating a different way of approaching interdisciplinary work across philosophy and the cognitive sciences.
2. Interdisciplinary philosophy: Mind and cognitive science

One problem that pervades philosophical theorizing in mind and cognitive science is that philosophical theories and distinctions can just simply fail to map onto (interesting) scientific distinctions. As illustrated below, even the apparently ‘good’ cases can often be seriously flawed. This problem at least shows that doing empirically informed philosophy is not as easy as it looks. On closer inspection of many cases where empirical work is used to support a philosophical theory, discussions of empirical research are either irrelevant to the philosophical theory or distinction, or they show it to be fundamentally misguided or uninteresting.

The second problem is that deriving a philosophical account of a mental phenomenon from empirical data, such as perception or consciousness, often (though not always) makes the philosophical account little more than a vague version of currently accepted scientific theories. Within science, vague theories are seen as inadequate theories. These philosophical accounts typically also attempt to provide a unified theory of wide explanatory scope, yet science tolerates very few such ‘grand unified’ theories. If philosophers aim to contribute to the mind/brain sciences by developing unifying but (largely) non-predictive theories, there is a real question about how scientifically useful such work is.

It is suggested below that these problems are serious and fairly common ones within at least some approaches to interdisciplinary work across philosophy and the mind/brain sciences. Instead of throwing in the towel and either retreating to an armchair or the local cognitive science department, I later propose an alternative, and hopefully more immediately productive, philosophical way to engage with empirical work.

2.1 Empirically informed philosophy of mind
Here I briefly describe an example of the way that philosophical distinctions and concepts can simply fail to fit scientific distinctions and concepts, even though empirical work is called upon to support the philosophical distinction. This example is taken from an area of research, and from the work of particular philosophers, that are typically seen as providing examples of good ways to engage with empirical work. The problems outlined below show just how careful one needs to be when calling on empirical work to support a philosophical theory. This example concerns the equation, recently made by several philosophers, between the contents of phenomenal consciousness and the contents of iconic (sensory) memory.

The general debate in which this example features is on whether (phenomenal) consciousness has rich or sparse content. It appears as though we have richly detailed experiences of whole scenes, but experimental work in the last ten years or so has thrown doubts on this. Work on inattentional blindness and change blindness (e.g. Mack and Rock, 1998, Simons and Rensink, 2005) showed that subjects failed to report significant changes in scenes (dancing gorillas, changes in conversational partners). Given our strong intuitions that we would notice a dancing gorilla, or a change in the identity of a conversational partner, the finding that we (typically) do not notice these changes raises doubts about how reliable these intuitions are. If we do not notice what we (intuitively) think we would notice, then perhaps we do not in fact consciously perceive as much as we think we do.

However, several philosophers have proposed alternative ways to interpret these experiments. It has been suggested that even if subjects do not overtly notice (or report) these changes or salient features, they may nevertheless be conscious of them. For example, Block (2007) suggests that rich phenomenal content is just not all concurrently reportable, Dretske (2007) distinguishes between knowing or thinking (hence reporting) about visual stimuli and consciously perceiving them, and Wolfe (1999) suggests that scenes could be consciously perceived in a richly detailed way but then quickly forgotten.
All these authors agree that a lack of overt report by subjects does not imply that they were not conscious of a scene in rich detail. The link between different types of access to information, or the relation between content and depth of processing, makes claims like these open to empirical evidence.

To provide empirical evidence for the claim that rich conscious content is experienced but is quickly ‘forgotten’, and is not all immediately accessible, many authors have turned to Sperling’s (1960) experiments on partial report superiority. In these experiments, subjects are shown a display of 12 letters for a very short duration, and then cued to report the contents of a particular row of letters. Subjects are able to report this row reasonably well, and the cued row can be anywhere on the display. However, subjects can only accurately report 3 or 4 letters (or the contents of one row) on each trial. That is, subjects have the potential, after viewing the display, or reporting almost all of the 12 letters, but can only actually access (report) 3 or 4 letters per trial.

This short term visual memory (iconic or sensory memory) of larger amounts of information that can be concurrently reported has been claimed by a range of philosophers to provide the rich contents of phenomenal consciousness. That is, the rich contents of phenomenal consciousness have been equated with the rich contents of sensory memory (e.g. Tye (2006), Block (2007), Fodor (2007, 2008) and Jacob and de Vignemont (2010)), thus providing support for the claim that consciousness has rich, rather than sparse, content.

However, a closer look at the details of current theories of sensory memory show that such an equation cannot hold. Instead of providing a rich store of visually detailed (non-conceptual) content that is consciously experienced, the type of short term visual memory investigated by Sperling is now recognized to be a non-unified memory store of differentially processed and decaying information (Loftus & Irwin, 1998, Luck & Hollingworth, 2008). As explained below, this memory store does not have the properties
that are attributed to phenomenal consciousness, but instead have very different properties that are not obviously realized in visual experiences. Further, there are alternative ways of explaining reports (and possibly intuitions) about the experience of rich conscious content. The claim that the contents of sensory memory are the contents of phenomenal consciousness is simply not consistent with contemporary scientific theories of sensory memory.

The main problem with philosophical accounts making these claims is that they assume sensory memory to be a single, unified phenomena relating to short-term storage of (non-conceptual) visual information that is present in consciousness. In contrast, contemporary theories use sensory memory as an umbrella term to refer to two distinct phenomena, where only one of which is directly related to visual experience. This phenomenon is not the one found in Sperling’s paradigm. Further, the type of sensory memory investigated by Sperling, and identified with phenomenal consciousness in the philosophical literature, displays rather different properties to those typically attributed to phenomenal consciousness.

One phenomenon found under the banner of sensory memory is visible persistence. This phenomenon is experienced as after-images, or the integration of spatially overlapping, but temporally separated patterns (patterns as seen ‘as one’ despite being presented at different times). Visible persistence is due to neural activity in early visual areas that continues when stimuli with very short display times, or high contrast stimuli (e.g. lightening in a dark sky), are no longer actually present. This is the phenomenon that is clearly related to visual experiences, and how long they persist over time (after-images can last for several seconds).

The other phenomenon found within sensory memory, and illustrated through Sperling’s work, is informational persistence. This refers to the short-term storage of visual information. Informational persistence consists of both a visible analogue store for shape
and location information (stored for 150-300ms), and a post-categorical store for abstract information such as identity (stored for 500ms). The differences in the type of information found in these two stores, and the different decay rates of this information, leads to specific kind of errors in the Sperling paradigm. When subjects are cued after the visible analogue store has decayed (location information gone), but while the post-categorical store is still available (identity information still available), subjects make ‘location errors’. In these cases, subjects correctly identify some letters, but since location information is no longer available, the letters come from non-cued rows. Perhaps of most importance, informational persistence is a type of visual memory; it tells us about the short-term storage of different types of visual information, but not about the content of duration of visual experiences. As Luck and Hollingworth (2008) state: “the partial-report technique does not measure directly the visible aspect of visual sensory memory, but rather that information persists after stimulus onset” (p. 16, original italics).

These facts pose several problems for accounts that equate the contents of rich phenomenal consciousness with the contents of sensory memory. First, several authors have treated visible and information persistence as different aspects of the same (conscious) phenomenon (e.g. Block, 2007, pp. 488-491; Tye, 2006, pp. 511-513), but these are misleading conflations of two very different phenomena. Visible and informational persistence are investigated with different experimental paradigms, concern different parts of the visual system, have different temporal properties, and relate either to early visual processing and visual experience, or to short-term visual memory. The Sperling paradigm, and the general phenomenon of informational persistence, are not measures of what is experienced, only what is ‘remembered’ and reported when the display is no longer present. Using informational persistence (memory) to make claims about the contents of visual experience is not accepted within current psychological theories, so neither should it be in philosophical accounts.

Second, even if we could make inferences from informational persistence to the contents
of phenomenal consciousness, the properties of the visible analogue and post-categorical stores are not consistent with the properties typically attributed to phenomenal consciousness. Information in these stores is deeply processed, up to the level of object identity. Subjects are not therefore reading off the letter identities from a conscious perception of detailed but (non-conceptual), visually detailed letter-like shapes. They are simply reporting the letter identities that are already processed and stored; this type of sensory memory is certainly not ‘iconic’. Further, subjects do not report a changing experience as different types of information degrade, as one might expect if the contents of sensory memory are the contents of phenomenal consciousness. Experience of letter displays do not suddenly lose spatial information, becoming a sort of non-spatial letter-identity soup, as one would imagine to be the case when subjects are making ‘location errors’ as described above.

Finally, the reports that subjects give of experiencing the whole display in a richly detailed way, can be explained in another, very different way. Instead of these reports being based on rich (internal) conscious content, expectation (de Gardelle 2009, Kouider et al., 2010) and scene gist (Oliva, 2005) are sufficient for reports (and the intuition) of phenomenal richness. They are sufficient even when the reported visual details are not in fact present; false reports of specific details based on the processing of scene gist are widely found and reliably generated (see e.g. Castelhano & Henderson, 2008). This means that the philosophical equation between the contents of sensory memory and phenomenal consciousness is inconsistent with contemporary psychological theories. Further, there is also an alternative explanation for the basis of reports (and likely intuitions) of phenomenal richness that is not based on the processing or storage of richly detailed (non-conceptual) visual information (for more details and potential implications, see Irvine, 2011).

This example illustrates how philosophical accounts that appear to deeply engage with empirical work can get it wrong. In scientific work, the details count, and can make the
difference between a claim being entirely consistent with empirical work or being entirely incompatible with it. Further, in this example (and others) philosophical work is often based on outdated scientific theories, or misinterpretations of current theories. Referring to historical philosophical theses, or alternative readings of philosophical theses, are viable moves in philosophy to motivate and justify a contemporary position. However, historical scientific theories cannot be used in this way, and much care needs to be taken in providing alternative interpretations of empirical evidence. If empirical work is used in philosophy, it needs to be used in a scientifically respectable way.

Of course, this does not show that philosophical discussions of conscious content are ultimately mistaken, misleading or uninteresting, to be abandoned entirely and replaced by psychological and neuroscientific theory. The point is that if philosophy of mind is to be properly informed by empirical work, then we will have to accept that this cannot just consist of searching for empirical support for philosophical theses wherever we can find it. Real interdisciplinary work across philosophy and science requires a detailed understanding of both disciplines in their own right, and attempting some degree of integration. This often results in the revision, sometimes deep revision, of the theories, concepts and distinctions we use across the disciplines in question. In this case, we will have to accept that some philosophical positions will turn out to be mistaken or uninteresting when confronted with contemporary scientific work.

Again, if we are committed to interdisciplinary work (and this seems to be a growing trend), then research methods in empirically informed philosophy may need to change in the ways outlined above. However, getting the balance right between scientific and philosophical contributions is difficult; an alternative problem that arises in empirically based philosophy is discussed below.

---

1 This is true across the board, see e.g. Kessel et al. (2008) on interdisciplinary research across health and the social sciences; McCauley and Bechtel (2001) and Churchland (1993) on cross-level research in psychology and into philosophy.
2.2 Deriving philosophical accounts from empirical work

This section concerns strongly empirically informed philosophy of mind and some parts of philosophy of cognitive science, where philosophical theories are derived from current scientific theories. That is, current empirical work is well-understood and well summarized, but more or less presented as philosophical work in its own right, or used to derive philosophical conclusions. There is no strong distinction between these cases and those addressed above, as being more familiar with empirical work does not rule out drawing faulty links between scientific and philosophical theories. However, more obviously basing philosophical accounts on empirical work faces additional problems. This problem is defining the role of the philosopher within cognitive science.

One option is to simply do (more or less) theoretical cognitive science, with one foot in cognitive science and one foot in the philosophy, sometimes being involved in experimental work, but more often offering interpretations and theoretical accounts of current scientific findings. I see no problem with philosophers essentially functioning as cognitive scientists engaged in theoretical work (we are supposed to be adaptable folk), but there are two related problems with this kind of approach.

One tactic in this kind of interdisciplinary work is to largely dispense with philosophical terms or to change them in such a way to fit experimental data. However, the resulting philosophical theory, based on scientific work, can be just a restatement of currently accepted theories in cognitive science, sometimes using different labels for different processes (or again, sometimes not). One example of this is Prinz’s (2012) theory of consciousness, which is largely just a recapitulation of attention based theories of consciousness, which have been popular in cognitive science for a long time (e.g. Baars, 1997, Dehaene & Naccache, 2001). Similar things were said about O’Regan and Noe’s (2001) sensori-motor account of consciousness; but in this case the claim was that while sensori-motor accounts are interesting, they are not new (see e.g. Gibson, 1979), and they
don’t necessary tell us much about what constitutes consciousness (e.g. Hardcastle, 2001). In providing an empirical account of what a particular phenomenon is (e.g. consciousness), these accounts do not add significantly to the existing scientific literature.

A more general, but related problem, is just what status empirically based philosophical theories have, particularly if we think that philosophical theories are (or should be) somewhat different from theories found in cognitive science. Again, there is no real problem in accepting that philosophers can just do theoretical cognitive science, but if they are to be seen as doing something that philosophers in particular do well, or if philosophical theorizing in cognitive science is to feed back into traditional philosophical questions (and it seems that this is the goal of many empirically-based philosophers), then it helps to identify what this is. One common approach is discussed here (though there may be others).

Philosophical theories that are based on empirical work often aim to provide a new way of thinking about a phenomenon (consciousness, perception) by considering a set of examples or a description of a mechanism, and situating this within some more general framework. This often includes making claims about what constitutes that phenomenon, identifying its function, and relating it to other mental phenomena (e.g. action, attention).

Of course such accounts can also be proposed by non-philosophers. Computational theories of perception and cognition also provide similar accounts; symbolic cognitive architectures like ACT-R make different claims about what cognition is compared with connectionist models, which are different again to probabilistic models (see e.g. Anderson & Lebiere, 2003; McClelland et al., 2003; McClelland et al., 2010; Griffiths et al., 2010). These different modeling frameworks give rise to very different claims about the nature of attention, perception, cognition, action, and more ‘philosophical’ notions such as representation and rationality (e.g. Gigerenzer et al. 1999).
Unfortunately, what seems to distinguish ‘philosophical’ theories of empirical phenomena is their enduring status as inadequate scientific theories; of being vague, qualitative, not specifying the boundary conditions of the theory, not generating predictions, and so on. These kind of theories are of course found in science, particularly when novel phenomena are found or when a theory has just been proposed. Another common feature of philosophical theories of empirical phenomena; they are often of the form of ‘grand unifying’ theories that attempt to account for a very wide range of phenomena (e.g. the theories of consciousness noted above). When on a par with other inadequate scientific theories, philosophical theories can therefore play a small but useful role in science; suggesting new frameworks to account for (and sometimes unify) a range of phenomena. However, a proliferation of these kind of theories is not necessarily very helpful.

First, there are very few successful theories in science that unify and explain a wide range of phenomena, but fail to exhibit the typical properties of ‘good’ theories. The theory of natural selection is a standard example; its explanatory scope is massive, but without the addition of context-specific knowledge and assumptions, it cannot make (quantitative) predictions, and the boundaries of the theory are still being worked out for specific cases (e.g. what proportion of evolution is really due to natural selection). As these kind of non-predictive, ‘grand unifying’ theories do not conform to the standard properties of scientific theories, very few of them are tolerated for long, or seen as useful explanatory frameworks in the first place. For example, Friston’s ‘free energy’ principle (e.g. Friston & Stephan, 2007; Friston et al., 2006) is potentially a very powerful ‘grand unifying’ theory for cognition and animal behavior in general. However, it appears to only be taken seriously in relevant research communities when empirical support is given for details of its implementation in specific cases (e.g. see Clark, 2012 (forthcoming) for a thorough overview of recent work, with much the same viewpoint). On a smaller scale, there is often complaint about the proliferation of models providing a ‘proof of concept’ for a particular explanatory framework, (that a model or framework can account for certain core cases, but does not provide predictions about any other cases), as these models also
often lack the properties of ‘good' scientific model (e.g. in decision-making see Glöckner & Betsch, 2011).

Philosophers aiming to give ‘grand unifying’ theories of mental/cognitive phenomena need to mindful of these facts. While there does seem to be a shortage of theoretical work in cognitive science (a gap philosophers could try to fill), the scientific value of adding more loosely specified, wide-scope theories into the literature is not obvious. ‘Good’ theoretical work needs to be highly empirically sensitive, and theories need to be good scientific theories; at the very least being able to explain and predict novel phenomena. If philosophers are joining in the game of interpreting experiments and posing grand unifying explanatory frameworks, then in order for these frameworks to be useful (or perhaps even considered) in cognitive science, they may end up having to do cognitive science anyway. This is likely to include formalizing their models, generating testable predictions, identifying critical properties (core predictions that could falsify the theory), and running experiments (see also Dennett, 2009, on this kind of enterprise).

2.3 Summary

I have not argued that all empirically informed or based philosophical accounts face the problem of working with inappropriate theoretical distinctions, or of providing nothing more than inadequate scientific theories, but just that within my experience these seem to be common problems facing empirically informed philosophy of mind and philosophy of cognitive science. In part, I suspect that this is because much of (interdisciplinary) philosophy of mind and cognitive science is focused on answering constitutive questions; what is perception, or attention, or consciousness, what are their functions, and how do they relate to other mental/cognitive/psychological states. Given the differences between philosophy of mind and contemporary cognitive science, it is hardly surprising that philosophical distinctions do not always map onto scientific ones, forcing many philosophers to either force a fit (often without realizing it), treat (constitutive)
philosophical questions as ones to be answered *a priori*, or fall into the bit of cognitive science that we are best trained to do (interpreting experimental results into new-ish conceptual frameworks, but not engaging in full-on scientific research).

The problems outlined above are not irresolvable, but I suggest below that there are other ways of pursuing interdisciplinary work that are as yet underutilized, but potentially invaluable. These are based on the idea of treating mental/cognitive/psychological properties as any other high-level property of a complex biological system, and trying to figure out the best way to investigate them. Rather than asking ‘what is (philosophically interesting phenomenon) X’, or ‘does X have (philosophically interesting) property Y’, we can instead ask ‘what methods can we use to investigate X (and whether X has property Y)’, and perhaps learn some surprising things in the process, both about how science works and about the phenomena we’re interested in. While a philosophy of science based approach is sometimes found in philosophy of cognitive science, I suggest that it deserves to be center stage.

3. An Alternative: Philosophy of the Science of the Mind

The kind of philosophy of science discussed below focuses on scientific methods and what these methods tell us about the phenomena we are interested in, along with apparently *a priori* metaphysical questions, such as what causation is (Woodward, 2008), what reduction means (Machamer & Sullivan, 2001), and what the difference is between correlation and identity (Bechtel & McCauley, 1999; McCauley & Bechtel, 2001). I will not argue for this approach in general here (see Bechtel, 2009, for more general examples of the utility of this approach), but illustrate, through two examples, the ways that insights from philosophy of science stand to affect both cognitive science and philosophy of mind. These are the status of first person data, and the way that we pose and answer constitutive questions (here, whether or not cognition can be extended).
3.1 First person data in science and philosophy

One discussion that stands to benefit from methodological insights from philosophy of science is the use of first person data (including introspection) in psychology and cognitive science. First person data is often treated as a unique way to access private data, and sometimes held (whether explicitly or not) to provide incorrigible evidence about the nature of a subject’s internal (conscious) states. While rather innocuous in some settings, such as patients reporting pain levels to medical practitioners, the veracity of first person data has become a talking point in consciousness studies. For example, Schwitzgebel (2008), while pointing out the surprising unreliability of at least some types of first person reports (our own emotional and bodily states, that there is little consensus over where there is any cognitive phenomenology), suggests that in some cases we can be trained to be better introspectors (2003), and that some methods of ascertaining first person data (e.g. buzzer method, 2007) might be better than others. Rather less carefully, the new field of neurophenomenology takes the contents of first person reports to be essential to studying consciousness, and that given suitable training and response scales, first person data is reliable and essential data (Ramsøy & Overgaard, 2004; Overgaard et al., 2006).

Despite the widespread use of first-person reports within science and philosophy, and the varying degrees of reliability ascribed to them, to date there has been relatively little discussion of methodological questions surrounding the use of first person data from the perspective of philosophy of science (though see Feest, 2012; Piccinini, 2005, 2009a, 2009b, 2010; Irvine, forthcoming). The following section outlines a number of questions about first person data that deserve wider discussion, and that may impact related areas in philosophy of mind. These questions call on the details of scientific measurement, knowledge about the mechanisms that generate first person reports, and possibly some rather difficult philosophical work on what first-person data really tells us about.
First, there is the essential question of whether the use of ‘private’ data can constitute a scientific method. Arguably, science must rely on (public) measures that can be verified by other means, and whose reliability we have some way of gauging. If first-person methods really are the only way of accessing the information they claim to inform us about, it is an open question what (if any) judgments we can make about their reliability. One way of solving this problem is to reject the assumption that first-person data is data about ‘private’ states, and see first-person data as public data, and a form of self-measurement (Piccinini, see esp. 2009a, 2009b, and Dennett’s heterophenomenology, 2003, 2007).

Yet even if first-person data is public like any other scientific data, there is the further question of how we interpret it. It is widely acknowledged that at least sometimes, reports are biased or inaccurate, but it is also widely assumed in the philosophical literature that there are cases where reports can be essentially bias free read-outs of an internal mental state (e.g. Bayne & Spener, 2010). However, as has long been known from the application of signal detection theory to human subjects (Green and Swets, 1966): “there is no such thing as an unmediated ‘subjective report’ – ever” (p. 526) Snodgrass and Lepisto (2007). ‘Free’ first person reports (no forced guessing) are the product of filtering a systems objective ability to respond to stimuli through a variable, context-sensitive threshold (criterion).

For example, the phenomenon of perceptual defense (e.g. Bruner & Postman, 1947, 1949) was used to support Freudian ideas about the unconscious by using an experimental set-up that implicitly varied subjects’ response bias. It was found that subjects were better at consciously perceiving (freely reporting) non-threatening words (‘shot’) than threatening words (‘shit’) when both sets of words were presented under identical conditions. This was taken as evidence of an internal ‘censor’ protecting the conscious mind from unpleasant stimuli. Instead, subsequent experiments and analysis using signal detection theory suggested that subjects perceive both sets of words, but that
college undergrads in the 1950s were rather wary of reporting swear words to experimenters, giving them a strong response bias against reporting the threatening words.

Although this is an extreme case, it shows how subjective reports can be affected by factors inherent in a task (e.g. inclusion of swear words), and factors relating to subjects (e.g. their motivation). The application of signal detection theory to human subjects showed that subjective reports are always, ineliminably, biased read-outs of the ‘objective’ ability of a system to respond to stimuli. In this case, eliminating a response bias is impossible if we want to sustain ‘free’ first reports. Minimizing a bias just means shifting the response towards an objective measure of sensitivity. Training subjects to use particular response scales is just to give them new response biases. How to use and interpret first person data is therefore something that demands discussion.

If first person reports are biased read-outs of ‘objective’ states then we seem to have a few options. First, we can use first person data as a direct means of characterizing the contents of mental states. This is the default but easy option; it is entirely possible that what a subject reports is different to what he thinks/feels/sees. Second, we can identify mental states with the objective facts about the system, and abandon ‘subjective’ data altogether. However, taking a reading of an ‘objective’ state of the system seems to have little to do with the original use of first person data; i.e. to investigate internal, subjective states of the person, not just the perceptual sensitivity of the visual system (see e.g. Lau, 2006, further discussion). Third, we can accept that we (currently) have little idea how first person data maps onto mental states, but try to make some head-way in particular cases, perhaps using comparisons to other types of data such as task training, motivation, expectation, knowledge about cognitive structure, and so on.

Yet Feest (2012) has also argued that in order to interpret introspective reports, we need to have some theory of how introspection works (see also Schwitzgebel 2010 on a
pluralist conception of the mechanisms of introspection). This is similar to the way that we need to know how a measuring tool works in order to interpret its readings (Hacking, 1983). Although first-person data is a reliable enough data in everyday life, if we are to use it to decide substantive scientific and philosophical questions (e.g. what are the contents/levels/types of consciousness, the boundaries of perceptual abilities), then we need to have a greater understanding of it as scientific data. In this case, just as we need to understand how a microscope works in order to interpret the data it provides, we need to understand how first-person data is generated in order to interpret it.

This relates to questions about what first-person data is actually data of. One key assumption that often underlies the use of first person data, and talk about how to interpret it, is the assumption that there really is a fact of the matter about what mental state a subject is in at a particular point in time. Given this assumption, and the problems identified above, subjects are now sometimes given training on how to figure out what they really are experiencing, and how best to report their mental states (e.g. Overgaard et al., 2006). However, there are other ways of thinking about what first-person data are data about. Dennett (2003, 2005, 2007) claims that first person reports tell us about a person’s beliefs about their experiences or mental states. In contrast, Piccinini (2010) suggests that they can instead be seen as direct reports about their experiences or mental states.

The problem, as noted above, is that we are very unsure how to map cognitive processes and their products (a range of behaviours, including reports) to mental states. Perhaps mental states just are whatever we think they are, or perhaps there are facts of the matter about what mental states we have at any one time. If the latter is true, we don’t know how much of a grasp of our mental life we really have (Schwitzgebel), and even if we did have a good knowledge of it, this doesn’t necessarily translate into first person data being faithful reflection of our mental states (e.g. due to response bias). What the relation is between first-person data and mental states, and the implications this has on what we
think mental states are, are questions that are central to the way we interpret first-person data, and deserve further serious discussion.

The aim of this section was to show that the methodological questions associated with the use of first person methods, and learning how first person reports and behaviours are generated and thus what they actually measure, are all more serious than is often given credit. More work from philosophy of science and cognitive science is needed to see how to interpret first person data. This includes further investigation of just which mechanisms generate the full range of first-person data (across ‘free’ reports, forced verbal report, button presses, etc), the essential properties of a scientific measure or scientific method of investigation, and how far problems with measuring and investigating mental states using first person data affects what questions we can ask about them.

The approach outlined above favours an investigation of fundamental questions about scientific measurement and the interpretation of data that underlie any scientific enterprise. This focus on general methodological questions, rather than on issues that are specific to particular research areas is a definite advantage to this approach. Instead of having to respond to long-standing intuitions and assumptions (e.g. first-person reports are largely correct, direct read-outs of private mental states), a focus on scientific methodology offers a new and potentially insightful approach towards this important range of questions. A philosophy of science approach is not the only way into these questions, but it does highlight many of them as essential methodological questions that are often bypassed, or are simply not part of the current dialectic.

3.2 Dichotomies, mechanisms and pluralism

A very different area where philosophy of science offers a new perspective is discussed in this section. This focuses on debates in philosophy of mind and cognitive science about
constitutive questions. The example considered here is the debate over whether cognitive processes only occur in the head, or whether they are (or can be) extended into non-neural bodily and environmental processes. The central problem here is the assumption from many philosophers that such questions must have a definitive yes or no answer, or more recently, that empirical investigation will provide a definitive yes or no answer. I suggest that taking inspiration from related discussions in philosophy of biology and neuroscience, we need not assume that constitutive questions have clear-cut answers, both in scientific practice, and in metaphysical terms too.

Clark and Chalmers (1998) suggested that cognition need not be limited to the processes that go on inside the skull. Instead, parts of cognitive processes could extend into external objects, such the physical manipulation of objects to solve spatial problems, or supplementary memory in the form of external devices. In order for cognition to remain a useful concept though, some conditions need to constrain what can count as cognitive extension (e.g. so that all the contents of the internet don’t count as external memories). Various criteria have been offered of what should count as an external item or process being part of a cognitive process, often based on a notion specific to cognition. In order to demarcate the conditions on cognitive extension, Clark and Chalmers (1998) appeal to the external resource being trusted and available, summed up in the parity principle: “If, as we confront some task, a part of the world functions as a process which, were it done in the head, we would have no hesitation in recognizing as part of the cognitive process, then that part of the world is (so we claim) part of the cognitive process” (p. 8, original italics). However, there has not been sufficient discussion of how similar the function must be, some conditions rely on questionable or vague definitions of cognition, and others simply don’t apply to all cases (e.g. Haugeland, 1998; Rupert, 2004; Wheeler, 2010; Adams & Aizawa, 2001, 2008).

As recently suggested by Kaplan (forthcoming), if the debate over cognitive extension comes down to identifying constitutive parts of a (cognitive) process, a problem often
faced in science, then seeing how this problem is routinely solved by scientists could help. Again, philosophy of science seems to have useful things to offer to debates in philosophy of mind and cognitive science, but I suggest that it does not quite provide the answer (or indeed the question) that Kaplan thinks it does.

In recent mechanistic philosophy of science, scientific explanation consists of identifying the multi-level entities and activities that, organized in particular ways, together produce the phenomenon of interest (Machamer at al., 2000). Craver (2007a, 2007b) has outlined the ways in which mechanisms are experimentally demarcated (i.e. how they are constituted), using the notion of mutual manipulability. The basic idea is that a part is a constitutive component of a mechanism (or cognitive process) if, by wiggling the part, the phenomenon wiggles, and by wiggling the phenomenon, the part wiggles. These mutual wiggles must be reasonably subtle and specific, in order to rule out background conditions from being included. Kaplan suggests that this generic way of identifying constitutive parts of a mechanism can be used to empirically establish whether or not a cognitive process is extended.

However, as Craver himself has argued (2009), mechanisms and their demarcation cannot be used to identify natural kinds or context-independent constitutive parts of a mechanism. What counts as being constitutively relevant for a mechanism in one research context may not count in another. The mutual wiggles must be reasonably subtle and specific, but how much so is up to the researcher and his/her research goal. If a therapist for Parkinson’s patients needs to find external ways of intervening on a patient to help with their memory loss, then a notepad may well count as a constitutive part of the cognitive process of memory, but if a neuroscientist is trying to create a drug to prevent neural plaque build up, then the use of a notepad may not be relevant at all.

To be sure, there are cases often cited in the extended cognition literature where it makes sense to include external objects or relations to external features as a constitutive part of the mechanism/cognitive process in question, because these mechanisms afford greater
predictive and explanatory power (for several compelling examples, see Kaplan, forthcoming; Chemero & Silberstein, 2008). But mechanisms and their range of explanatory power do not in themselves determine the definitive boundaries of a cognitive process; they only specify a reasonable way to carve up a system given a description of a phenomenon, and a set of research goals. In this case, considerations about which fineness of grain to use, strength of coupling, and so on, come right back into the picture of how to carve up the mechanism:

“The boundaries of mechanisms…cannot be delimited by spatial propinquity, compartmentalization, and causal interaction alone. This is because the spatial and causal boundaries of mechanisms depend on the epistemologically prior delineation of relevance boundaries. But relevance to what? The answer is: relevance to the phenomena that we seek to predict, explain, and control…the mechanistic structure of the world depends in part upon our explanatory interests and our descriptive choices…” (Craver, 2009, pp. 590-591)

So far, this only seems to affect how we talk about cognition, not what it really is. In this way of thinking, empirical questions have no import for metaphysical questions. Yet similar work in philosophy of biology, particularly related to complex multi-level systems, suggests that just as there are multiple ways of carving up mechanisms, there are multiple ways of carving the same process up into ‘real’ ontological chunks. Ontological pluralism, accepting that there are multiple and equally viable ways of carving up the same bit of reality into ontological types or kinds, is not necessary a radical position, but one that naturally arises from the consideration of scientific methods (see e.g. Boyd, 1999; Wilson, 2005; Craver, 2009; Dupré, 1993).

While this obviously flies in the face of so much debate about this issue, it seems that ideas from philosophy of science, properly applied, have much potential to calm similar debates within philosophy of mind and cognitive science. This is because, as has become
clear in biology, and now reflected in philosophy of biology, complex biological systems are not the sort of thing that support exception-free generalisations or clear-cut dichotomies. Neither is it profitable to model biological systems in only one way, or carve up a causal landscape in only one way. Reflection on biology itself has generated philosophical positions of experimental, cross-level, conceptual and theoretical pluralism because these are the best ways we have of understanding biological systems (e.g. Kellert et al., 2006).

These are also reflected in metaphysical positions about biological systems; the current science is structured in a pluralist way not just because we aren’t very good at it, but because sharp, generalizable, context-independent boundaries are just not a feature of biological systems – they are not something we can ever expect to find. We, as evolved cognizing beings, are clearly complex biological systems, so insights from philosophy of biology have obvious application in philosophy of mind and cognitive science. Future projects would include seeing how far similar tactics work in other debates (e.g. related topics of embodied cognition, whether cognition involves representations, dual systems theories vs. sensori-motor theories of perception, and so on).

4. Conclusions

I have suggested that at least some contemporary approaches to interdisciplinary work across philosophy and the mind/brain sciences face some serious problems. These are recognizing and dealing with the mis-fit between philosophical and scientific distinctions (possibly making deep revisions necessary in philosophical accounts), and how to establish a useful role for philosophers in cognitive science that does not result in overpopulating the field with vague theoretical frameworks. Not all interdisciplinary work falls into these traps, these problems do not rule out these approaches as being productive, and these are not the only ways of characterizing the ways in which
philosophers engage with empirical work (see e.g. Brook, 2009; Dennett, 2009; Thagard 2009).

Instead, I have suggested that there are alternative approaches that are currently underused and may provide insights that are not easily reached using other methods. Approaches from current philosophy of science offer a more direct way to engage with scientific methods, what they tell us about specific phenomena, and how scientific methodology can inform philosophical questions about mental phenomena. Aside from the two case studies introduced above, there are many other general debates in philosophy of mind and cognitive science that could be invigorated, clarified or laid to rest by looking to the counterpart discussions in philosophy of science: these include discussions of causation, reduction, explanation, emergence, cross-level identity claims, and the individuation of concepts and scientific/natural kinds.

While ideas from philosophy of science that offer to defuse or radically re-shape debates in philosophy of mind and cognitive science may not be very palatable for those currently embroiled in them, this is not, I think, a reason to ignore these ideas. While not committal about whether “philosophy of science is philosophy enough” (Quine, 1953, p. 446), I hope to have shown that it certainly deserves far greater attention than it currently receives in empirically informed philosophy of mind and cognitive science.

References


Beauchamps, K. (1999). Constructing a philosophy of science of cognitive science. *Topics in...


26


McClelland, J. L., Botvinick, M. M., Noelle, D. C., Plaut, D. C., Rogers, T. T., Seidenberg, M. S., & Smith, L. B. (2010). Letting structure emerge: connectionist and dynamical systems approaches to cognition. *Trends in Cognitive Sciences*, 14(8), 348-356.


Piccinini, G. (2009a). Scientific methods ought to be public, and Descriptive Experience Sampling is one of them. *Journal of Consciousness Studies, 18*(1), 1-12.


