Philosophy of Psychology
Edouard Machery

Abstract: Philosophy of psychology takes various forms. Some philosophers of psychology use psychological findings and theories to develop new answers to traditional philosophical issues. A smaller number of philosophers of psychology take their cue from the philosophy of science. They describe and evaluate the discovery heuristics, theories, and explanatory practices endorsed by psychologists. Finally, much philosophy of psychology can be characterized as psychological theorizing. Just like psychologists, philosophers propose empirical theories of specific aspects of our mind, trying to explain relevant psychological phenomena. Focusing mostly on this aspect of the philosophy of psychology, I will consider philosophers' contribution to the theoretical development of psychology in four areas: cognitive architecture and modularity (§2); situated, embodied and extended cognition (§3); concepts (§4), and mindreading (§6).[1] Before doing this, however, I will discuss philosophers' and psychologists' views and arguments about the distinctive character of psychology—its mentalistic nature (§1).

1. The Scientific Legitimacy of Mentalism?

1.1 The Place of Mental States in Psychological Theories and Explanations

It will be useful to start the discussion of the place of mental states in psychology with an example. Everyday experience and experimental evidence suggest that people often reason poorly about probabilistic matters. For instance, Tversky and Kahneman asked participants to read the following story:

> Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations (1982, 92).

Participants were then asked to rank various "statements by their probabilities," including the following three:

(1) Linda is active in the feminist movement
(2) Linda is a bank teller
(3) Linda is a bank teller and is active in the feminist movement

Remember that the conjunction axiom of probability theory states that the probability of a conjunction is always smaller or equal than the probability of one of its conjuncts:

(4) $P(p\&q) \leq P(p)$

Thus, participants in Tversky and Kahneman's experiment would be mistaken to answer that it is more probable that Linda is a feminist and a bank teller than simply a bank teller. Nonetheless, 89% of participants judged that Linda was more likely a feminist bank teller than a bank teller, a mistake known as "the conjunction fallacy."

---

[1] Philosophers have contributed to the theoretical development of other areas, such as emotions (e.g., (Griffiths, 1997); (Prinz, 2004a)), consciousness (e.g., (Block, 1995), Block, forthcoming; Chalmers, 1996, (Chalmers, 2004); (Noë, 2004)), perception (e.g., (Jacob & Jeannerod, 2003); (Noë, 2004)) psychopathology (e.g., Murphy, 2007), moral psychology (e.g., Doris, 2002; (Nichols, 2004); Doris & Stich, 2006; Prinz, 2007), the relation between language and thought (e.g., (Carruthers, 2006)), and the scientific value of evolutionary psychology (e.g., (Buller, 2005); (Machery & Barrett, 2006); (Machery, Forthcoming-a)). For the sake of space, I will not review these contributions here.

Tversky and Kahneman use such mistakes to investigate the psychological mechanisms underlying (correct and incorrect) probabilistic judgments. They propose that people's probabilistic judgments result from simple psychological processes (called "heuristics") that often lead to correct judgments, but occasionally mislead (they are then called "biases")—hence the name of their research program, "heuristics and biases." Thus, according to Tversky and Kahneman, people often use a simple psychological process, called "the representativeness heuristic," to make probability judgments. People evaluate the probability that *a* is an F, according to the similarity between the description of *a* and the stereotype of an F. In the experiment just described, people evaluate the probability that Linda is a bank teller by comparing the description of Linda that is provided in the cover story and the stereotype of a bank teller. Because Linda is less representative of a bank teller than of a feminist bank teller, people rank (3) as more probable than (2), thereby committing the conjunction fallacy.

For present purposes, what matters is that Tversky and Kahneman's account of people's probabilistic judgments is *mentalistic*: it involves ascribing *mental states* to people (viz., internal states that mediate between environmental stimuli and behavior[2]) and *psychological processes (*viz., processes that manipulate mental states). Consider again the representativeness heuristic. When people evaluate the probability that an individual *a* is an F, they retrieve a stereotype of an F (a mental state) from memory. This stereotype is compared with the information about *a*, a psychological process that results in a measure of how representative *a* is of Fs. The probability that *a* is an F is a monotonic function of this measure. Mentalism (viz., the appeal to mental states, psychological processes, and other psychological entities such as personality traits) is a characteristic property of the theories and explanations developed in the various subfields of psychology (e.g., social psychology, cognitive psychology, personality psychology, etc.).

Now, one might wonder whether mentalist theories are legitimate scientific theories. Mental states, psychological processes, and other psychological entities are unobservable entities, which are posited to account for behavior. Like other theoretical entities, claims to their reality should be subject to scrutiny. More importantly, mental states and psychological processes have often been associated with ontological and epistemological properties that are not scientifically kosher. Since the seventeenth century, mental states have often been associated with substance dualism—the idea that there are two substances, matter and mind. But, if mental states were distinct from physical states, it would be mysterious how they could causally interact with physical states. Furthermore, it has sometimes been argued that by introspection (the observation of one's own mental states), each of us has a privileged access to one's own mental states. But, this first-person privilege seems at odds with the idea that evidence in science is public and accessible from a third-person perspective.

1.2 Methodological Behaviorism

In the first decades of the 20th century, the school of psychology, known as "behaviorism" or "methodological behaviorism,"[3] formulated the most radical answer to the question "Can psychological entities be legitimately postulated by a scientific theory?" For behaviorists, mental states and other psychological entities had no place in psychology. Behaviorists not only contended that referring to unobservable states between environment and behavior was not required for explaining behavior, but they also argued that it was

[2] Because they are diverse, defining what mental states are is a difficult task. Many mental states have semantic properties: they can be true or false (this is the case of, e.g., beliefs) or satisfied or unsatisfied (this is the case of, e.g., desires). Some mental states also have phenomenal properties: it feels something to have them.

[3] "Methodological" is used to distinguish the type of behaviorism discussed here from Ryle's (1951) logical behaviorism, according to which mental state predicates pick out behavioral dispositions.

unscientific. In his influential behaviorist manifesto, John Watson, the father of behaviorism, wrote that psychology could be written without ever using "the terms consciousness, mental states, mind, content, introspectively verifiable, imagery, and the like… It can be done in terms of stimulus and response, in terms of habit formation, habit integrations and the like" (1913, 166-167).

Behaviorism was a reaction against the dominant psychology of the time.[4] Much of human psychology in the second half of the 19th century was based on introspection. By the end of the 19th century, however, the nature of introspection and its value as a scientific method had become a controversial topic among introspective psychologists themselves (Caldwell, 1898; Titchener, 1899). By contrast, following the lead of Edward Thorndike and Robert Yerkes, animal psychology had developed controlled and reproducible experimental designs and quantitative measuring techniques that allowed for the experimental study of numerous animal behaviors (e.g., orientation, problem solving, etc.). The disarray of introspective psychology and the successes of animal psychology paved the way for the reception of behaviorism.

For Watson (1913), the rejection of mental states and other psychological entities as objects of scientific study and as scientific explanatory entities was primarily due to his rejection of introspection. In substance, Watson argues that because introspection is not a proper scientific methodology, the states to which it gives access (viz., the mental states) have no place in a scientific psychology. We will come back to this argument later.

Introspection itself was rejected on the grounds that its products were subjective and unreliable. Watson argued that far from being objective observational reports, introspective reports were influenced by psychologists' theoretical commitments. He also noted that introspection had failed to promote any consensus among psychologists. In addition, introspection prevented the unification of psychology, since it was not used in animal psychology. Rejecting introspection could allow for the transfer of methods from animal to human psychology and for the comparison of results across disciplines.

In addition, one finds in Watson's manifesto the following parsimony argument. Because, for Watson, (animal and human) behavior is an instinctive (i.e., inherited) or habitual (i.e., learned) reaction to measurable aspects of the environment, explaining, predicting, and manipulating behavior merely requires knowledge of the learning history of the agent and of its environment. Introspective data about mental states have no role to play for explaining, predicting, and manipulating behavior.

Although distinct behaviorist theories have been developed (e.g., by Clark Hull, B.F. Skinner, and Edward Tolman), these theories shared a common focus on the contingencies between behaviors (called "responses") and measurable environmental conditions (called "stimuli"). Behaviorists attempted to explain why specific behaviors were produced in specific environments by looking at the history of interactions between organisms and their environment. They developed two main accounts of learning: classical conditioning and operant conditioning. Importantly, none of these accounts is mentalist: no reference is made to intervening variables between behavior and the environment.

According to the theory of classical conditioning, inspired by Pavlov's work, organisms have spontaneous responses ("unconditioned responses") caused by environmental stimuli ("unconditioned stimuli"). When a stimulus that is not associated with any response (a "conditioned stimulus") is repeatedly presented in association with an unconditioned stimulus, the organism ends up associating the response with the conditioned stimulus. Thus, in Pavlov's well-know experiment, a sound was repeatedly played when food was presented to a dog (causing the dog to salivate) and the dog ended up salivating at the mere hearing of this sound.

The theory of operant or instrumental conditioning, developed by Thorndike and Skinner (e.g., 1938), divides behaviors into two types: responses, which are caused by

[4] Watson was also influenced by the work of the Russian physiologist Ivan Pavlov. See O'Donnell (1985) for a history of behaviorism.

identified stimuli (e.g., salivating when food is present), and operants, which are not associated with specific stimuli (e.g., pressing a lever for a rat). Focusing on operants, Skinner proposed that organisms tend to repeat operants whose strength is "reinforced."[5] Thus, if a rat receives some food when it presses a lever, the rat will tend to press the lever again. The operant is reinforced and the strength of the operant is measured by how long the organism will press the lever at a rate higher than the base rate (i.e., the rate before reinforcement) under extinction (that is, when no reinforcer follows the operant).

While classical conditioning can explain why an organism (a human or an animal) extends a behavior that is already part of its behavioral repertoire to new contexts, operant conditioning can explain the inclusion of new, originally randomly produced behaviors in the behavioral repertoire of an organism.

Behaviorism has certainly had a lasting and, in many respects, positive influence on psychology. Modern psychology inherited its emphasis on controlled experimental procedures and quantitative, objective measures (rather than introspective reports). Classical and operant conditioning are also important properties of learning (but see Gallistel & Gibbons, 2001).

Still, behaviorism has been rejected in the second half of the twentieth century for four main reasons.[6] First, the explanatory scope of behaviorist theories turned out to be limited. In his influential review of Skinner's *Verbal Behavior*, Noam Chomsky (1959) noted that while the key notions of stimulus, reinforcer, and operant were well defined, when they were applied to pigeons and rats whose behavior and environment are highly constrained by Skinnerian experimental designs (such as a Skinner box), they were poorly defined outside such a context. He concluded that when used to explain everyday (human and animal) behavior (e.g., language acquisition by children), Skinner's theoretical notions were either misapplied or a misleading paraphrase of mentalistic notions. Second, psychologists and philosophers have come to realize that to explain people's behavioral competences, it is necessary to postulate intervening states, which mediate between environment and behavior. Thus, Chomsky (1959) argued that it was impossible to explain language acquisition without considering both the environment in which learning takes place (the linguistic stimuli) and the contribution of the learner. Third, we saw above that the rejection of mentalism was principally a consequence of the rejection of introspection. Just like behaviorists, contemporary psychologists typically deny that introspection is a valuable source of evidence about mental states and psychological processes. However, in contrast to behaviorists, they do not conclude that mental states and psychological processes are not proper explanatory entities and objects of scientific study. For contemporary psychologists, the ascription of mental states in psychological explanations is to be justified on explanatory grounds (viz., to account for behavioral competences) rather than on the basis of introspection. As a result, the unreliability and subjectivity of introspective reports do not impugn the justification of mental state ascription. Finally, as we shall see in the next section, philosophers and psychologists developed a new characterization of mental states and psychological processes that made mentalism scientifically reputable.

1.3 The Computational Representational Theory of Mind

The limits of behaviorism show that a purely behavioral psychology is unpromising and that internal states have to be postulated to account for behavioral competences. At this juncture, it seems natural to propose that mental states are the internal states needed for a

---

[5] The definition of reinforcement was a subject of controversy among behaviorists. I will overlook this difficulty here.

[6] The history of the demise of behaviorism is more complex that is typically acknowledged. Behaviorism remains influential in animal psychology, in educational psychology, and in some fields of neuroscience. Furthermore, although the lore has it that Chomsky's scathing review of Skinner's *Verbal Behavior* was the crucial event in the rejection of behaviorism, efforts to reevaluate mentalism were already on their way in the 1940s (e.g., MacCorquodale & Meehl, 1948).

scientific psychology. This would be premature, however, for it remains to provide a scientifically satisfying account of mental states and psychological processes.

Information theory and the theory of digital computers have provided such an account ((Fodor, 1975); (Newell & Simon, 1976); (Pylyshyn, 1984); (Marcus, 2001); see Bechtel and Herschbach's essay in this volume for additional detail). Mental states are thought to be representations, that is, particulars endowed with a specific content. Written or spoken sentences, maps, paintings, and road signs are representations in that sense: they represent the world as being so and so and can be thereby true or false, accurate or inaccurate. Sentences, maps, paintings, etc., have a derived content, meaning that they have a given content because people use them to represent in a given way. By contrast, mental representations have a non-derived or original content, because, on pain of regress, they cannot intentionally be used to represent. Two mental states of a given type (e.g., the beliefs that Paris is in France and that Berlin is in Germany) are distinguished by the content of their respective representations. Different types of mental states (e.g., beliefs and desires) are distinguished by their functional roles. While the belief that it is noon and the desire that it is noon both are representations that express the proposition that it is noon, they are distinguished by their functional role: beliefs and desires have different causal connections with perceptual stimuli, other mental states, and actions.

Psychological processes consist in transformations of representations. Philosophers and psychologists have proposed that these transformations are computational—hence the name "the computational representational theory of mind." That is, in substance, these transformations are governed by rules that apply to representations in virtue of their formal properties. These rules do not apply to representations in virtue of their content, but in virtue of some non-semantic properties of the representations, in exactly the same way as numerals do not get added by pocket calculators in virtue of their meaning (the numbers they express), but in virtue of their syntactic properties (see Piccinini, 2008 for complications).

The computational representational theory of mind assuages worries about the scientific legitimacy of mentalism. Digital computers show that material entities can implement computational processes manipulating representations. They also illustrate how being introspectively accessible is not an essential property of representations. Still, it is important to flag two issues raised by this theory. First, what is the relation between mental representations and brain states? Second, in virtue of what do mental representations have their content? Philosophers of mind and of psychology have extensively discussed these two issues, but for the sake of space, I will not elaborate on them in this article (see, e.g., (Fodor, 1987); Stich & Warfield, 1994).

2. Cognitive Architecture and Massive Modularity

The organization of the processes that underwrite our perceptual and cognitive architecture—"the cognitive architecture"—is an important topic of debate among philosophers of psychology. Jerry Fodor (1983) has proposed an influential hypothesis about the nature of human cognitive architecture. He distinguishes two types of processes, modules and non-modular processes. A Fodorian module is a psychological process that has most of the following properties: it has a specific type of inputs and it produces shallow or non-conceptual outputs; its functioning is fast, automatic, cognitively impenetrable (that is, other systems have no access to and no influence upon its internal processing), and informationally encapsulated (that is, it has access to only a subset of the information that is represented in the mind); it is also realized in a discrete brain area, it is innate, and it breaks down in characteristic ways.[7] By contrast, non-modular processes have few (if any) of these properties. For Fodor, modules underwrite a few capacities—particularly, our perceptual capacities and our linguistic faculty. The processes underlying our higher cognitive capacities

[7] In *The Mind Does not Work that Way* (2000), Fodor emphasizes the encapsulation of modules.

(e.g., the fixation of our beliefs, the determination of our desires)—what Fodor calls "our central processes"—are supposed to be non-modular.

In contrast to Fodor, many psychologists have argued that the processes underlying some higher cognitive capacities are modular. For instance, Elisabeth Spelke has argued that our capacity to orient ourselves is underwritten by a geometric module (Hermer & Spelke, 1996), while Nancy Kanwisher has proposed that a module underwrites our capacity to identify individual faces (Kanwisher, McDermott, & Chun, 1997). Going further, some psychologists (e.g., John Tooby, Randy Gallistel) and some philosophers (e.g., Peter Carruthers, Dan Sperber) propose that *all* our psychological processes are modular, a thesis known as "the massive modularity hypothesis."

Various arguments have been proposed in support of the massive modularity hypothesis (Sperber, 1994, 2001; for a systematic overview, see Carruthers, 2006, Chapter 1). I focus here on evolutionary psychologists John Tooby and Leda Cosmides's argument that evolution is unlikely to have selected for non-modular psychological processes ((Tooby & Cosmides, 1992); Cosmides & Tooby, 1994).[8] Rather than focusing on the properties that are characteristic of Fodorian modules (see above), Tooby and Cosmides characterize modules in terms of functional specialization (sometimes called "domain-specificity"): modules have been selected for bringing about a specific outcome (that is their function). Evolutionary modules contrast with "domain-general" processes, viz. psychological processes that are not functionally specialized. Tooby and Cosmides assume (as I will for the sake of the argument) that psychological processes are adaptations, that is, (using evolutionary psychologists' terminology) traits that have been selected because they solved some adaptive problems (e.g., finding food, choosing a fertile mate, avoiding poisons, detecting cheaters, etc.). They argue that a domain-general process would be less efficient than a modular process to solve a given problem, for the latter, but not the former, would have been designed to solve this problem. Thus, natural selection would tend to favor modular processes over non-modular processes. As Cosmides and Tooby famously put it (1994, 89), "as a rule, when two adaptive problems have solutions that are incompatible or simply different, a single general solution will be inferior to two specialized solutions. In such cases, a jack of all trades is necessarily master of none, because generality can be achieved only by sacrificing effectiveness."

In reply, Richard Samuels (1998) notes that there are two types of modules—computational modules and Chomskyan modules. Computational modules are mechanisms; they are defined by the nature of their processes. The modules hypothesized by Tooby and Cosmides are of this first kind. Chomskyan modules are bodies of knowledge about specific tasks—they are representations, not processes. Chomskyan modules can be used by domain-general reasoning mechanisms. To illustrate this contrast, consider the adaptive problem of avoiding poisonous foods. (Because their diet is not specialized, omnivores have had to solve this problem.) A computational module for solving this problem would be a mechanism for distinguishing safe from unsafe foods. By contrast, a Chomskyan module would be a body of knowledge about safe and unsafe foods, which could be used by a domain-general reasoning mechanism. Having distinguished these two types of module, Samuels notes that natural selection would not prefer a cognitive system made of computational modules to a cognitive system made of Chomskyan modules used by a domain-general reasoning system, because Chomskyan and computational modules are equally specialized for solving adaptive problems. But, if the mind were a cognitive system made of Chomskyan modules used by a domain-general reasoning system, the massive modularity hypothesis would be false. Thus, Samuels concludes, it does not follow from the hypothesis that our cognitive architecture is the product of evolution by natural selection that the mind is massively modular.

The massive modularity hypothesis has been criticized on various grounds ((Fodor, 2000); Buller, 2005). I discuss here only two problems, the input problem (Fodor, 2000) and

[8] Some psychologists and philosophers have also proposed tractability arguments, which state that only modular processes can perform the computations that are required by the tasks defining our cognitive and perceptual capacities (for review and discussion, see (Samuels, 2005)).

the brain evolution problem ((Quartz, 2002)). Noting that a specific type of inputs is required to trigger a given module, Fodor (2000) contends that a psychological process (a routing system) is needed to pair each module with the stimuli that trigger it. Because this routing system would have to be activated by all types of stimuli, it could not be modular. Thus, the massive modularity hypothesis is false. Clark Barrett (2005) has convincingly rebutted this argument by drawing an analogy between enzymes and modules. Enzymes come into contact with a large range of substrates. However, because they have specific binding sites, only some of these substrates are bound with enzymes. Similarly, modules could have access to all representations but be activated by only some of them. No non-modular routine process is thus needed in a modular mind.

Barbara Finlay and colleagues' work on brain evolution has also inspired an important objection against the massive modularity hypothesis. They found that across mammals, the volume of the main brain structures is correlated to the volume of the whole brain ((Finlay & Darlington, 1995)). Steve Quartz has argued that these findings show that natural selection did not act on individual brain structures independently of the other brain structures (2002, 189):

> These results, suggest that neural systems covary highly with one another as a consequence of the restricted range of permissible alterations that evolutionary psychology can act upon. This makes the massive modularity hypothesis of narrow evolutionary psychology untenable (2002, 189).

Quartz's argument should be resisted ((Machery, Forthcoming-c)). A closer look at Finlay and colleagues' data shows that across mammals, the volume of the whole brain does not covary perfectly with the volume of the main brain structures, suggesting that natural selection may have acted upon their volume. Furthermore, there is more to brain evolution than the volume of the brain structures considered by Finlay and colleagues. Natural selection probably acted upon the nature of brain cells, their organization, or the connectivity between brain areas.

Finally, it is noteworthy that a careless use of the term "module" has muddled the debate about the massive modularity hypothesis. "Module" means different things for different people (Barrett & Kurzban, 2006). Particularly, as noted above, evolutionary psychologists define modules as those processes that have a dedicated function. They need not have any of the properties that characterize Fodorian modules. (Similarly, "module" has a distinctive use in neuroscience.) For instance, modules need not be innate nor need they be automatic or cognitively impenetrable (Machery & Barrett, 2006). Rejecting evolutionary psychologists" massive modularity hypothesis on the grounds that our central processes do not possess the properties that characterize Fodorian modules (e.g., they are not automatic, etc.) is thus unsound.

3. Embodied, Situated, and Extended Cognition

Traditionally, philosophers and psychologists hold that the mind receives some information about its environment through the senses, uses this information to reason and make decisions, which may lead to action. Some philosophers and psychologists, whose views are often grouped together under the headings "embodied cognition," "situated cognition," and "extended cognition," have criticized this conception of the relation between cognition and the cognizer's environment. Although the views denoted by these headings differ in some respects, for simplicity, I will use the expression "extended cognition" in what follows, noting the differences between these views when appropriate. It has often been noted that this new movement combines several distinct positions without clearly marking their differences (e.g., (Wilson, 2002); Rupert, 2004).[9] In this section, I briefly distinguish four

[9] In addition to the works cited in this section, see Clark, 1997; Anderson, 2003; Wilson, 2004; (Shapiro, 2004).

threads, before discussing in some detail the idea that mental states and psychological processes are not located in the brain.

A first thread is methodological. Proponents of extended cognition contend that a proper understanding of psychological processes involves examining the environment in which cognition takes place (e.g., Hutchins, 1995)—a position often referred to by the label "situated cognition." This methodological claim is sometimes justified on the grounds that psychological processes are designed (by evolution or by learning) for specific (physical and social) environments. To illustrate, according to Gerd Gigerenzer and Ulrich Hoffrage ((Gigerenzer & Hoffrage, 1995)), the processes underlying probabilistic reasoning are designed to manipulate representations of natural frequencies, rather than probabilities, consistent with the fact that for most of human history, probabilistic information was only available in the form of natural frequencies.

A second thread highlights the importance of agency in understanding cognition (e.g., Noë, 2004; Gallagher, 2005). This emphasis is supposed to stand in contrast to cognitive psychology's traditional focus on situations that involve no or little action (e.g., chess playing, remembering words on a list, etc.).

A third, more radical thread takes issue with the idea that cognition involves manipulating representations (e.g., (Brooks, 1991); (Thelen & Smith, 1994)). Anti-representationalists typically focus on some phenomena that proponents of representation-based approaches to cognition explain (or would explain) by means of representations and representation-based processes. They then explain these phenomena without positing any process that manipulates representations. On this basis, they draw the following induction: if postulating representations is not needed to explain these phenomena, behavior and cognition at large can be explained without representations (for discussion, see, e.g., Vera & Simon, 1993; (Clark & Toribio, 1994)).

A fourth strand of argument, often referred to by the labels "extended cognition" and "extended mind," focuses on the location of mental states and psychological processes. Philosophers and psychologists have often identified token mental states with brain states and psychological processes with neural processes. In sharp contrast, Mark Rowlands (1999, 22) writes that "[c]ognitive processes are not located exclusively inside the skin of cognizing organisms," while Andy Clark and David Chalmers argue:

> [W]e will argue that *beliefs* can be constituted partly by features of the environment, when those features play the right sort of role in driving cognitive processes. If so, the mind extends into the world (1998, 12).

According to this view, at least some token mental states are external to the body or involve extra-corporeal objects as proper parts, while cognition involves the manipulation of these entities.

Two well-known examples might usefully illustrate this view. Consider first how we perform a complex arithmetical operation by hand, such as the multiplication of 37 by 23 (Clark & Chalmers, 1998; Adams & Aizawa, 2001, 2008; Noë, 2004). We write down one numeral below the other. Focusing on the rightmost digital of each numeral ("7" and "3"), we multiply the numbers they express. We write down "1" on a third line and write "2" as a carry-over. We then multiply 3 (the number expressed by the leftmost digital of the first numeral) by 3 (for the rightmost digital of the second numeral) and add the carry-over. We write down the numeral "11" left of the numeral "2" (and so on). To perform this multiplication, we create and manipulate objects (viz., numerals) that are external to our body in a rule-governed manner. According to proponents of the extended cognition, the numerals are part of our mental states and their rule-governed manipulation counts as psychological processing. As Alva Noë puts it (2004, 220), "[i]f the pencil and paper are necessary for the calculation, why not view them as part of the necessary substrate for the calculating activity?"

Consider a second example. Clark and Chalmers (1998) propose that in some situations, a notebook can literally be part of someone's memory. They compare a normal woman, Inga, who relies on her memory to determine the address of the Museum of Modern Art, with an Alzheimer patient, Otto, who relies on his constantly available notebook to determine the address of the museum. Clark and Chalmers contend that in spite of the differences between Inga's and Otto's cases, both Otto and Inga believe that the Museum of Modern Art is located on 53rd street:

> To provide substantial resistance, an opponent has to show that Otto's and Inga's cases differ in some important and relevant respect. But in what deep respect are the cases different? To make the case *solely* on the grounds that information is in the head in one case but not in the other would be to beg the question. If this difference is relevant to a difference in belief, it is surely not *primitively* relevant. To justify the different treatment, we must find some more basic difference between the two (1998, 6).

This last example is useful to bring to the fore the central argument for the view that mental states and psychological processing extend beyond the skin: there is no significant difference between some states that involve extra-corporeal entities and some brain states or between the manipulation of extra-corporeal entities, such as consulting a notebook, and the manipulation of mental representations, such as consulting one's memory. If there is really no significant difference between them, then some states that involve extra-corporeal objects are genuine mental states and some processes that involve manipulating these objects are genuine psychological processes.

Unsurprisingly, this fourth thread has caused a fair amount of discussion among philosophers. Most critics grant that if there were no significant differences between states of the brain and states involving extra-corporeal objects as proper parts or between processes involving only brain states and processes involving extra-corporeal objects, then not all mental states and psychological processes would be in the head, but they deny that the antecedent of this conditional is satisfied. Particularly, endorsing the computational representational theory of mind (§1), Fred Adams and Ken Aizawa (2001) have argued that mental states are representations that are endowed with an original content and that psychological processes are computational processes defined over these representations. For them, these two properties are "the mark of the cognitive." Because the extra-corporeal objects that are manipulated (for instance, the addresses in Otto's textbook) do not have any original content, they do not count as mental states and their manipulations do not count as psychological processes.

There are two main worries with Adams and Aizawa's argument. First, it rests on controversial (though widespread) necessary conditions for something to be a mental state and for something to be a psychological process, ones that might be rejected by proponents of extended cognition. Second, accepting Adams and Aizawa's necessary conditions, proponents of extended cognition might reply that a state count as mental provided that some of its parts have an original content and that a process counts as psychological, provided that some steps in this process involve states with original content (or states with some parts having an original content).

Robert Rupert's (2004) main argument against extended cognition does not fall prey to these worries, because he does not assume a specific mark of the cognitive. Rather, focusing on memory, he highlights the differences between the properties of memory retrieval on the one hand and the use of extra-corporeal objects to store information on the other. A large number of generalizations have been found about how people store information in memory (e.g., interference effects[10]) and how they retrieve information from memory (e.g., recency effects[11]). He correctly notes that few of these generalizations apply

[10] Associating two objects makes it more difficult to associate one of them with a new object.
[11] Objects recently memorized are easier to retrieve from memory.

to the gathering of information from physical mnemonic aids, such as notebooks. Furthermore, any generalization that could apply to information retrieval from both memory and mnemonic aids would probably be about a much larger class of systems, which would include, but not be identical to, the class of cognitive systems. Rupert concludes that treating states and processes within the brain and states and processes involving extra-bodily objects as physical parts is not a promising strategy for cognitive science.

4. Concepts

People classify objects intro classes, samples into substances, and events into event types. This capacity, typically called "categorization," is a basic capacity of human cognition: without it, we would be unable to acquire any general knowledge. Psychologists assume that when we categorize an object into a class (often called "a category" in psychology), we rely (maybe unconsciously) on some knowledge about this class.[12] Thus, when I classify an object as a table, I use some knowledge about tables. Psychologists call "concepts" those bodies of knowledge that are used by default to categorize (for an overview of the psychology of concepts, see (Murphy, 2002) and (Machery, Forthcoming-b); for a history of the psychology of concepts, see (Machery, 2007a)). Importantly, in addition to categorization, concepts are also used by the psychological processes underlying other capacities, such as induction. Thus, a concept of water is a body of knowledge about water that is used by default to categorize samples as being samples of water, to reason inductively about water, and so on. Psychologists interested in concepts attempt to describe the properties of these bodies of knowledge.

Philosophers have long paid attention to the psychology of concepts.[13] Famously, Fodor has argued that all the theories of concepts developed in psychology were incorrect ((Fodor, 1994), 1998). His favorite target has been the prototype theory of concepts ((Rosch & Mervis, 1975); (Hampton, 1979)). Prototype theorists argue that a concept is a body of statistical knowledge about a class (or a substance, etc.). The simplest versions propose that a prototype is a body of knowledge about the typical properties of the members of a class. Thus, a prototype of a dog is a body of knowledge about the typical properties of dogs. Fodor's main objection against the prototype theory can be put simply. Concepts compose. Thus, anybody who can think about dogs and about blue things can *ipso facto* think about blue dogs. But prototypes don't compose. Thus, concepts cannot be prototypes. To support the second premise of this argument, Fodor has put forward several considerations. The pet fish argument is the best known of these. Fodor notes that a poodle might be a prototypical pet and that a shark might be a prototypical fish, while a prototypical pet fish is a golden fish. Thus, our prototype of a pet fish is not derived from our prototype of a pet and our prototype of a fish. Rather, it is derived from our experience with pet fish. Thus, prototypes don't compose.

Fodor's pet fish argument is unconvincing. The fact that the prototype of a pet fish does not result from the combination of the prototype of a pet and of the prototype of a fish does not show that prototypes do not compose, *when we have no experience with members of the extension of the complex concepts*. That is, we might combine a prototype of an *x* (say, of a spy) and the prototype of a *y* (say, of a grandmother), when we have no experience with objects that are *x* and *y* (spy grandmothers). Experimental evidence does suggest that people produce complex prototypes in these conditions (see Murphy, 2002, Chapter 12 for review). Particularly, people seem to assume that the typical properties of an *x* and of a *y* tend to be also typical of an object that is both an *x* and a *y* ((Hampton, 1987)). Thus, a grandmother spy has the typical properties of a grandmother: she might have grey hair and wear out-of-fashion clothes (see, however, (Connolly, Fodor, Gleitman, & Gleitman, 2007)).

[12] Here, "knowledge" is used as psychologists do. It roughly means information or misinformation.

[13] In addition to the issues discussed here, philosophers have also discussed the acquisition of concepts ((Fodor, 1981); (Laurence & Margolis, 2002)).

Recent philosophical work on concepts has focused on two main issues: whether an empiricist theory of concepts is viable ((Prinz, 2002)) and whether concepts form a natural kind ((Machery, 2005)). I consider these two issues in turn. Following psychologist Lawrence Barsalou ((Barsalou, 1999)), Jesse Prinz (2002) has argued that recent developments in the psychology of concepts support a view of concepts that has many affinities with Hume's empiricist theory of ideas. Although there are several differences between Barsalou's, Prinz's, and others' neo-empiricist theories of concepts, they all endorse the two following theses ((Machery, 2006b)):

(1) The knowledge that is stored in a concept is encoded in several perceptual representational formats;
(2) Conceptual processing involves reenacting some perceptual states and manipulating these perceptual states.

(1) is about how we encode our conceptual knowledge (Prinz, 2002, 109). Neo-empiricists assume that each perceptual system involves a distinct representational format. (1) asserts that our conceptual knowledge is encoded in these perceptual representational formats. By contrast, amodal theorists argue that our conceptual knowledge is encoded in a representational format that is distinct from our perceptual representational formats. That is, for amodal theorists, we possess a distinct, *sui generis* representational format, which is used to encode our conceptual knowledge, in addition to our perceptual representational formats. This distinct representational format is usually thought of as being language-like. To illustrate this distinction, according to neo-empiricists, Marie's conceptual knowledge of apples consists of the visual, olfactive, tactile, somatosensory, and gustative representations of apples that are stored in her long-term memory. These representations are a subset of the perceptual representations of apples Marie has had in her life. According to amodal theorists, Marie's conceptual knowledge of apples consists of representations encoded in a single, distinct representational format.

(2) concerns the nature of the psychological processes underlying categorization, induction, deduction, analogy-making, linguistic comprehension, and so forth. The central insight is that retrieving a concept from long-term memory during reasoning, categorization, etc., consists in producing some perceptual representations. For instance, retrieving the concept of dog when we reason about dogs consists in producing some visual, auditory, etc., representations of dogs. This process is called "simulation" or "reenactment." Thinking about dogs during reasoning, thus, consists of simulating seeing, hearing, and smelling dogs. Our psychological processes consist in manipulating these reenacted percepts. Thus, according to Barsalou, when we decide whether some object has a given part, for example whether lions have a mane, we produce a visual representation of a lion and another of a mane and we match these two representations; if they do match, we decide that lions have a mane ((Solomon & Barsalou, 2001), 135-136).

Two main lines of reply to these neo-empiricist theories of concepts have been developed in the philosophical literature. Some philosophers and psychologists have argued that neo-empiricist theories of concepts suffer from the very same problems that plagued David Hume's theory of ideas. For instance, John Sarnecki ((Sarnecki, 2004)) and Arthur Markman and Hunt Stilwell ((Markman & Stilwell, 2004)) argue that Prinz's empiricist account of concepts cannot be applied to abstract concepts (see also Machery, 2006b), while Markman and Stilwell argue that it cannot be applied to relational concepts, such as the concept of an uncle (see (Prinz, 2004b) for a rejoinder to Sarnecki and Markman and Stilwell).

Instead of discussing the theoretical problems of the neo-empiricist theories of concepts, I have focused on the evidence for these theories, for it is supposed to favor neo-empiricism over amodal theories of concepts ((Machery, 2007b)). Thus, Barsalou and colleagues write:

> Amodal theories have been attractive theoretically because they implement important conceptual functions, such as the type-token distinction, categorical inference,

> productivity, and propositions. (…) Conversely, *indirect* empirical evidence has accumulated for modality-specific representations in working memory, long-term memory, language, and thought ((Barsalou, Simmons, Barbey, & Wilson, 2003), 85-86).

The evidence for neo-empiricism is not as strong as might appear at first. Three problems plague most of the current empirical research inspired by neo-empiricism. First, neo-empiricists erroneously assume that for a given experimental task, a single prediction can be derived on behalf of amodal theorists and can be tested experimentally. In fact, however, there are numerous competing amodal models of the psychological process involved in a given task and these competing models make different predictions about subjects' performance in this task. Because different amodal models make different predictions, it is not the case that Barsalou's and others' experimental findings are inconsistent with an amodal view of concepts in general. Rather, they are inconsistent with specific amodal models. Second, neo-empiricists have not acknowledged that amodal theorists, such as Fodor or Zenon Pylyshyn, recognize that we can use imagery to solve various problems. Thus, when one is asked to count the number of windows in one's own house, one typically visualizes one's house and counts the number of visualized windows. Because amodal theorists admit the existence and importance of imagery, they do predict that in some tasks people will simulate having perceptual states. But neo-empiricist researchers have often failed to focus on tasks for which amodal theorists would not expect people to use perceptual imagery. Finally, neo-empiricists have not acknowledged the possibility that in some domains, or for some tasks, or, maybe, in some contexts, people might use perceptual representations, while using amodal representations in other domains, or in other tasks, or in other contexts.

The second debate philosophers of psychology interested in concepts concerns whether the class of concepts is a natural kind. Most psychologists working on concepts assume that concepts share many scientifically important properties and attempt to describe those properties. They assume thereby that concepts constitute a natural kind, that is, roughly, a class of entities about which numerous non-accidental, scientifically important generalizations can be made, an assumption I called "the Natural Kind Assumption" (Machery, 2005, forthcoming). Against the Natural Kind Assumption, I have argued that most classes are represented by several concepts that belong to kinds that have little in common. For instance, I propose that the class of dogs is typically represented by several concepts of dog and that these concepts have few (scientifically relevant) properties. To support this proposal, I have shown that for each relevant cognitive capacity (categorization, induction, concept combination), some phenomena are best explained if one posits a first kind of concepts (namely, prototypes), other phenomena are best explained if one posits a second kind of concepts (what psychologists call "exemplars"), and yet other phenomena are best explained if one posits a third kind of concepts (what psychologists call "theories").[14] Because these three kinds of concepts have little in common, I have concluded that concepts are not a natural kind. Furthermore, I have proposed that the notion of concept is ill-suited for a scientific psychology and that the term "concept" should be eliminated from its theoretical vocabulary, exactly as the notion of superlunear objects was eliminated from astronomy.

The claims that concepts are not a natural kind and that the notion of concept should be eliminated from psychology have come under criticism. While agreeing that there are different kinds of concepts, Gualtiero Piccinini and Sam Scott ((Piccinini & Scott, 2006);

[14] In substance, an *exemplar* is a representation of an individual. Thus, an exemplar of dog is a representation of a particular dog. According to the exemplar view of concepts, a concept is a set of exemplars. In substance, according to the *theory* view of concepts, a concept is similar to a scientific theory. Thus, a concept of dog might consist of some nomological knowledge about dogs.

see (Machery, 2006a) for a reply) have argued that if most classes were represented by several concepts (for instance, if dogs were represented by a prototype, by a set of exemplars, and by a theory), then, contrary to the conclusion I drew, concepts would be a natural kind. For, each concept consists of several parts. For instance, the concept of dog would have three parts (one corresponding to the prototype of a dog, one corresponding to exemplars of particular dogs, one corresponding to a theory about dogs). Dan Weiskopf ((Weiskopf, Ms)) has developed a different criticism. While highlighting the diversity of concepts, he argues that numerous generalizations can in fact be made about concepts. He concludes that eliminating "concept" from the theoretical vocabulary of psychology would prevent the formulation of numerous important generalizations.

5. Mindreading

Mindreading is the practice of ascribing mental states, such as beliefs, desires, emotions, and perceptions, to others and to oneself. It is an essential and automatic component of our everyday life. Consider watching Alfred Hitchcock's *Rear Window*. At the end of the movie, L.B. Jeffries, played by James Stewart, hides in the dark in his apartment. We understand this behavior because we know that he *believes* that the killer is coming to his apartment and because he does not *want* to be killed. Stewart's behavior is meaningful because we have ascribed some specific mental states to the character he is playing.

Philosophers of psychology have been involved in an interdisciplinary attempt to characterize the psychological mechanisms underlying mindreading (an attempt that also involves developmental psychologists, psychopathologists, neuropsychologists, and animal psychologists). Two main accounts have been developed: the theory theory and the simulation theory. I consider them in turn.[15]

Although proponents of the theory theory, such as philosophers Peter Carruthers, Fodor, Shaun Nichols, and Steve Stich and psychologists Simon Baron Cohen, Alison Gopnik, Alan Leslie, and Joseph Perner, disagree on various points, they concur that people have a large and complex body of knowledge about mental states, the relations between mental states, the relations between mental states and stimuli, and the relation between mental states and behaviors.[16] For instance, we might know that if an individual sees that *p*, she typically believes that *p*. We might also know that for an individual to see an object, she has to stand in some physical relation to this object (e.g., her face has to be turned toward this object, her eyes have to be directed toward this object, etc.). People use this body of knowledge when they ascribe beliefs and desires to others:

> The central idea shared by all versions of the 'theory-theory' is that the processes underlying the production of most predictions, intentional descriptions, and intentional explanations of people's behavior exploit an internally represented body of information (or perhaps mis-information) about psychological processes and the ways in which they give rise to behavior (Stich & Nichols, 1995), 87-88).

Note that we need not be aware that we possess and use this body of knowledge: it might be subdoxastic, exactly as our linguistic knowledge is supposed to be.

Simulation theorists, such as philosophers Alvin Goldman, Robert Gordon, and Jane Heal, and neuropsychologists Jean Decety and Vittorio Gallese, doubt that we have an extensive body of knowledge about mental states and that mindreading involves using this body of knowledge. They contend that ascribing mental states to someone else consists in

[15] Philosophers have also participated to the debates about whether non-human primates can mindread (e.g., Povinelli & Vonk, 2004), about whether the theory of mind is modular (e.g., Fodor, 1992), and about the nature of mechanisms underlying self-knowledge (e.g., (Nichols & Stich, 2003); Carruthers, 2006).
[16] Wellman, 1990; Perner, 1991; Baron-Cohen, 1995; (Gopnik & Meltzoff, 1997).

simulating her mind, or, to put it metaphorically, to put oneself in her shoes.[17] Gallese and Goldman write:

> ST [simulation theory] arose partly from doubts about whether folk psychologizers really represent, even tacitly, the sorts of causal/explanatory laws that TT [theory theory] typically posits. ST suggests that attributors use their own mental mechanisms to calculate and predict the mental processes of others. (Gallese & Goldman, 1998, 496).

In spite of a few disagreements about the nature of simulation, simulation theorists agree that simulating involves mimicking the mental life of the individuals to whom we want to ascribe mental states. Mindreading involves having similar states and similar thought processes as the individuals to whom mental states are ascribed. For Goldman, for instance, simulation typically involves three steps. One first pretends to have some mental states. Thus, while playing chess, if one attempts to predict one's partner next move, one pretends to have the mental states one's partner might have (e.g., her desire and knowledge of chess). Second, these pretend states are used as inputs to one own's reasoning and decision processes. Once one has pretended to have some mental states, one reasons as if these mental states were one's own. Thereby, one mimics the chain of thoughts others might have. For instance, to predict the next move of one's chess partner, one decides what move to make, pretending to have the mental states one's chess partner might have. Finally, one ascribes a mental state to others. For instance, instead of acting on one's decision about what move to make, this decision is "taken off-line" and used to predict what one's partner will do.

Philosophers have spent much energy clarifying these two approaches and contrasting various versions of both approaches. Rather than looking at these details, I will now sketch some of the main arguments developed on behalf of each approach.

Stich and Nichols (1992) have developed an influential argument for the theory theory. They note that specific biases influence the processes involved in reasoning and in decision-making. They focus particularly on a phenomenon called "the endowment effect" (Thaler, 1980): people are only willing to sell an object that they possess for more money than they paid when they acquired it. For instance, people might be willing to pay $10 to acquire the poster of a movie, but be unwilling to sell it for less than $15. Now, consider a situation where we have to predict what price someone is going to ask for selling an object and what price she would be willing to pay for acquiring this object. Suppose, as Goldman would have it, that we use our own decision processes to make these predictions. Then, because our own decision processes are biased, we should predict that others will fall prey to the endowment effect. The theory theory is not committed to this prediction because it might not be part of our (maybe implicit) theory of mind that people's decision processes are so biased. Thus, Stich and Nichols write:

> If there is some quirk in the human decision making system, something quite unknown to most people that leads the system to behave in an unexpected way under certain circumstances, the accuracy of predictions based on simulations should not be adversely affected (1992, 263).

Stich and Nichols note that people are very poor at predicting that they themselves and others would be victims of the endowment effect.

Goldman and other proponents of simulation theories have replied that systematic errors in mindreading are in fact consistent with simulation theory (see, e.g., Goldman, 2006; for discussion, see (Stich & Nichols, 1995)). Simulation is accurate only when the pretend states (e.g., in a chess game, pretending to have one's partner's knowledge of chess and her desire to win the game) are accurate. If the pretend states differ in a systematic manner from the states that the target of the simulation actually has, simulation will lead to systematic

[17] Gordon, 1986; Heal, 1986; (Goldman, 1989), (Goldman, 1992); Meltzoff & Decety, 2003.

mistakes. In reply, a theory theorist might concede that simulation theory predicts systematic mistakes, while questioning whether it predicts the very mistakes highlighted by Stich and Nichols.

I now focus on some arguments for the simulation theory. In the 1990s, neuropsychologists Giacomo Rizzolatti, Vittorio Gallese, and colleagues discovered some neurons in the ventral premotor cortex of macaques (area F5) that not only fire when the macaques are doing an action (as was expected), but that also fire when the macaques observe another macaque or the human experimenter do the same action (Gallese, Fadiga, Fogassi, & Rizzolatti, 1996). These neurons have been dubbed "mirror neurons" (for recent reviews, see (Rizzolatti & Craighero, 2004); (Gallese, 2007)). Evidence suggests that humans also have mirror neurons. Remember that according to simulation theorists, mindreaders mimic others: they have the same (or similar) mental states and they go through the same (or similar) chains of thoughts. Gallese and Goldman have proposed that the existence of mirror neurons supports simulation theory, because they are activated both when one decides to act and when one understand others' actions (1998, 498):

> MN activity seems to be nature's way of getting the observer into the same 'mental shoes' as the target—exactly what the conjectured simulation heuristic aims to do (1998, 498).

Goldman has also argued that neuropsychological findings show that the recognition and ascription of emotions, such as fear, disgust, or anger, involves a simulation process (Goldman & Sripada, 2005; Goldman, 2006). Roughly, when another person expresses an emotion facially and behaviorally, we are supposed to experience the emotion that this person is experiencing. We then ascribe to her the emotion that we are experiencing (see Goldman & Sripada, 2005 for a careful discussion of several models of this process). Seeing John make a disgust face (brows narrowed, upper lip raised, lip corners drawn down and back, and nose drawn up and wrinkled) causes disgust in me. Because I recognize that I feel disgust, I ascribe disgust to John. Goldman and Sripada have convincingly argued that the hypothesis that emotion ascription involves a simulation process accounts for the finding that following lesions in the brain areas involved with specific emotions, patients who are unable to experience these emotions are also impaired in recognizing them.

Recent philosophical contributions to the understanding of mindreading have partly moved beyond the original debate between the simulation theory and the theory theory. Researchers are now developing various hybrid accounts of the mechanisms underlying mindreading—that is, accounts that include both simulation and theory-based psychological processes (Nichols & Stich, 2003; Goldman, 2006).

6. Conclusion and Future Directions

In this essay, I have reviewed some important issues that have been at the center of the philosophy of psychology: the legitimacy of mentalism, the modularity of the cognitive architecture, the situated and extended nature of cognition, the nature of concepts, and the mechanisms underlying mindreading. To conclude, I will sketch what the philosophy of psychology may possibly look like in the forthcoming years. Certainly, philosophers will probably continue contributing to the areas discussed in this article (see footnote 1 for other areas of active debate), but they will also hopefully turn to new issues. These might include questions about the methodology of psychology, which has so far attracted little attention among philosophers (but see Trout, 1998; Glymour, 2001), such as the role of null hypothesis significance testing (Trout, 1999; Fidler, 2005; Machery, ms) and the relation between group data and the study of the mind of individuals. In addition, philosophers of psychology will probably attempt to improve our understanding of the relation between psychology and neuroscience (e.g., Schouten & Louren de Jong, 2007).[18]

References

[18] I would like to thank Luc Faucher and my research assistant Kara Cohen for their comments on a previous version of this article.

Adams, F., & Aizawa, K. (2001). The bounds of cognition. *Philosophical Psychology*, *14*, 43-64.
Adams, F., & Aizawa, K. (2008). *The Bounds of Cognition*. Oxford: Blackwell.
Anderson, M. L. (2003). Embodied cognition: A field guide. *Artificial Intelligence*, *149*(1), 91-103.
Baron-Cohen, S. (1995). *Mindblindness: An Essay on Autism and Theory of Mind*. Cambridge, MA: MIT Press.
Barrett, H. C. (2005). Enzymatic computation and cognitive modularity. *Mind and Language, 20,* 259-287.
Barrett, H. C., & Kurzban, R. (2006). Modularity in cognition: Framing the debate. *Psychological Review, 113,* 628-647.
Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*, 577-660.
Barsalou, L. W., Simmons, W. K., Barbey, A. K., & Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Sciences, 7*, 84-91.
Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences, 18*(2), 227-287.
Block, N. (Forthcoming). Consciousness, accessibility and the mesh between psychology and neuroscience. *Behavioral and Brain Sciences*.
Brooks, R. A. (1991). Intelligence without Representation. *Artificial Intelligence, 47*, 139-159.
Buller, D. J. (2005). *Adapting Minds: Evolutionary Psychology and the Persistent Quest for Human Nature.* Cambridge, MA: MIT Press.
Caldwell, W. (1899). The postulates of a structural psychology. *Psychological Review, 6*, 187-191.
Carruthers, P. (2006). *The Architecture of the Mind*. New York: Oxford University Press.
Chalmers, D. J. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford University Press.
Chalmers, D. J. (2004). How can we construct a science of consciousness? In M. S. Gazzaniga (Ed.), *The Cognitive Neurosciences III* (pp. 1111-1120). Cambridge, MA: MIT Press.
Clark, A. (1997). *Being There: Putting, Brain, Body and World Together again*. Cambridge, MA: MIT Press.
Clark, A., & Chalmers, D. J. (1998). The extended mind. *Analysis, 58*, 7-19.
Clark, A., & Toribio, J. (1994). Doing without representing? *Synthese, 101*, 401-431.
Connolly, A. C., Fodor, J. A., Gleitman, L. R., & Gleitman, H. (2007). Why stereotypes don't even make good defaults. *Cognition, 103*, 1-22.
Cosmides, L., & Tooby, J. (1994). Origins of domain specificity: The evolution of functional organization. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the Mind: Domain Specificity in Cognition and Culture* (pp. 85-116). Cambridge: Cambridge University Press.
Doris, J. (2002). *Lack of Character: Personality and Moral Behavior.* New York: Cambridge University Press.
Doris, J., & Stich, S. (2006) Moral psychology: Empirical approaches. In E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy*, http://plato.stanford.edu/entries/moral-psych-emp/ (Retrieved 02/01/2008).
Fidler, F. (2005). *From Statistical Significance to Effect Estimation: Statistical Reform in Psychology, Medicine and Ecology*. PhD Dissertation, Department of History and Philosophy of Science, University of Melbourne.
Finlay, B. L., & Darlington, R. B. (1995). Linked regularities in the development and evolution of mammalian brains. *Science, 268*, 1578-1584.
Fodor, J. A. (1975). *The Language of Thought.* Cambridge, MA: Harvard University Press.

Fodor, J. A. (1981). The present status of the innateness controversy. In J. A. Fodor, *Representations: Philosophical Essays on the Foundations of Cognitive Science* (pp. 257-316). Cambridge, MA: MIT Press.

Fodor, J. A. (1983). *The Modularity of Mind: An Essay on Faculty Psychology*. Cambridge, MA: MIT Press.

Fodor, J. A. (1987). *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge, MA: MIT Press.

Fodor, J. A. (1992). A theory of the child's theory of mind. *Cognition*, *44*, 283-296.

Fodor, J. A. (1994). Concepts: A potboiler. *Cognition, 50*, 95-113.

Fodor, J. A. (1998). *Concepts: Where Cognitive Science Went Wrong*. Oxford: Oxford University Press.

Fodor, J. A. (2000). *The Mind Doesn't Work that Way*. Cambridge, MA: MIT Press.

Gallagher, S. (2005). *How the Body Shapes the Mind*. Oxford: Oxford University Press.

Gallese, V. (2007). Before and below "theory of mind": Embodied simulation and the neural correlates of social cognition. *Philosophical Transactions of the Royal Society B: Biological Sciences, 362*, 659-669.

Gallese, V., & Goldman, A. I. (1998). Mirror neurons and the simulation theory of mindreading. *Trends in Cognitive Sciences, 2*(1), 493-501.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, *119*, 593-609.

Gallistel, C. R., & Gibbon, J. (2001). Computational versus associative models of simple conditioning. *Current Directions in Psychological Science*, *10*, 146-150.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review, 102*, 684-704.

Glymour, C. (2001). *The Mind's Arrows: Bayes Nets and Graphical Causal Models*. Cambridge, MA: MIT Press.

Goldman, A. I. (1989). Interpretation psychologized. *Mind & Language, 4*, 161-185.

Goldman, A. I. (1992). In defense of the simulation theory. *Mind & Language, 7*, 104-199.

Goldman, A. I. (2006). *Simulating Minds*. New York: Oxford University Press.

Goldman, A. I., & Sripada, C. (2005). Simulationist models of face-based emotion recognition. *Cognition*, *94*, 193-213.

Gopnik, A., & Meltzoff, A. N. (1997). *Words, Thoughts, and Theories*. Cambridge, MA: MIT Press.

Gordon, R. (1986). Folk psychology as simulation. *Mind & Language*, *1*, 158-171.

Griffiths, P. E. (1997). *What Emotions Really Are: The Problem of Psychological Categories*. Chicago: University of Chicago Press.

Hampton, J. A. (1979). Polymorphous concepts in semantic memory. *Journal of Verbal Learning and Verbal behavior, 18*, 441-461.

Hampton, J. A. (1987). Inheritance of attributes in natural concept conjunctions. *Memory & Cognition, 15*, 55-71.

Heal, J. (1986). Replication and functionalism. In J. Butterfield (Ed.), *Language, Mind and Logic* (pp. 135-150). Cambridge: Cambridge University Press.

Hermer, L., & Spelke, E. S. (1996). Modularity and development: The case of spatial reorientation. *Cognition*, *61*, 195-232.

Hutchins, E. (1995). *Cognition in the Wild*. Cambridge, MA: MIT Press.

Jacob, P., & Jeannerod, M. (2003). *Ways of Seeing, the Scope and Limits of Visual Cognition*. Oxford: Oxford University Press.

Tversky, A., & Kahneman, D. (1982). Judgments of and by representativeness. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment Under Uncertainty: Heuristics and Biases* (pp. 84-98). Cambridge: Cambridge University Press.

Kanwisher, N., McDermott, J., & Chun, M.M. (1997): The fusiform face area: A module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience*, *11*, 4302-4311.

Laurence, S., & Margolis, E. (2002). Radical concept nativism. *Cognition, 86*(1), 22-55.

MacCorquodale, K., & Meehl, P. E. (1948). On a distinction between hypothetical constructs and intervening variables. *Psychological Review, 55*, 95-107.
Machery, E. (2005). Concepts are not a natural kind. *Philosophy of Science, 72*, 444-467.
Machery, E. (2006a). How to split concepts. Reply to Piccinini and Scott. *Philosophy of Science, 73*, 410-418.
Machery, E. (2006b). Two dogmas of neo-empiricism. *Philosophy Compass, 1*(4), 398-412.
Machery, E. (2007a). 100 years of psychology of concepts: The theoretical notion of concept and its operationalization. *Studies in History and Philosophy of Biological and Biomedical Sciences, 38*, 63-84.
Machery, E. (2007b). Concept empiricism: A methodological critique. *Cognition, 104*, 19-46.
Machery, E. (Forthcoming-a). Discovery and confirmation in evolutionary psychology. In J. J. Prinz (Ed.), *Oxford Handbook of Philosophy of Psychology*. Oxford: Oxford University Press.
Machery, E. (Forthcoming-b). *Doing Without Concepts*. New York: Oxford University Press.
Machery, E. (Forthcoming-c). Massive modularity and brain evolution. *Philosophy of Science*.
Machery, E. Ms. How null hypothesis significance testing obstructs progress in psychology.
Machery, E., & Barrett, H. C. (2006). Debunking *Adapting Minds*. *Philosophy of Science, 73*, 232-246.
Marcus, G. F. (2001). *The Algebraic Mind*. Cambridge, MA: MIT Press.
Markman, A., & Stilwell, H. (2004). Concepts á la modal: An extended review of Prinz's Furnishing the mind *Philosophical Psychology, 17*(3), 391-401.
Meltzoff, A. N., & Decety, J. (2003). What imitation tells us about social cognition: A rapprochement between developmental psychology and cognitive neuroscience. *Philosophical Transactions of the Royal Society: Biological Sciences, 358*, 491-500.
Murphy, D. (2007). *Psychiatry and the Scientific Image*. Cambridge, MA: MIT Press.
Murphy, G. L. (2002). *The Big Book of Concepts*. Cambridge, MA: MIT Press.
Newell, A., & Simon, H. A. (1976). Computer science as empirical enquiry. *Communications of the Association for Computing Machinery, 19*, 113-126.
Nichols, S. (2004). *Sentimental Rules. On the Natural Foundations of Moral Judgment*. New York: Oxford University Press.
Nichols, S., & Stich, S. (2003). *Mindreading: An Integrated Account of Pretense, Self-awareness and Understanding Other Minds*. Oxford: Oxford University Press.
Noë, A. (2004). *Action in Perception*. Cambridge, MA: MIT Press.
O'Donnell, J. M. (1985). *The Origins of Behaviorism: American Psychology, 1870-1920*. New York: New York University Press.
Perner, J. (1991). *Understanding the Representational Mind*. Cambridge, MA: MIT Press.
Piccinini, G. (2008). Computation without representation. *Philosophical Studies*, *137*, 205-241.
Piccinini, G., & Scott, S. (2006). Splitting concepts. *Philosophy of Science, 73*(4), 390-409.
Povinelli, D.J., & Vonk, J. (2004). We don’t need a microscope to explore the chimpanzee mind. *Mind & Language*, *19*, 1-28.
Prinz, J. J. (2002). *Furnishing the Mind: Concepts and their Perceptual Basis*. Cambridge, MA: MIT Press.
Prinz, J. J. (2004a). *Gut Reactions: A Perceptual Theory of Emotions*. New York: Oxford University Press.
Prinz, J. J. (2004b). Sensible ideas: A reply to Sarnecki and Markman and Stilwell. *Philosophical Psychology, 17*(3), 419-430.
Prinz, J. J. (2007). *The Emotional Construction of Morals*. New York: Oxford University Press.
Pylyshyn, Z. W. (1984). *Computation and Cognition*. Cambridge, MA: MIT Press.
Quartz, S. R. (2002). Toward a developmental evolutionary psychology: Genes, development, and the evolution of the human cognitive architecture. In S. J. Scher &

F. Rauscher (Eds.), *Evolutionary Psychology: Alternative Approaches* (pp. 185-206). Dordrecht: Kluwer.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience, 27*, 169-192.

Rosch, E., & Mervis, C. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7*, 573-604.

Rowlands, M. (1999). *The Body in Mind: Understanding Cognitive Processes*. Cambridge: Cambridge University Press.

Rupert, R. (2004). Challenges to the hypothesis of extended cognition. *The Journal of Philosophy, 101*, 389-428.

Ryle, G. (1951). *The Concept of Mind*. London: Hutchinson.

Samuels, R. (1998). Evolutionary psychology and the massive modularity hypothesis. *The British Journal for the Philosophy of Science*, *49*, 575-602

Samuels, R. (2005). The complexity of cognition: Tractability arguments for massive modularity. In P. Carruthers, S. Laurence, & S. Stich (Eds.), *The Innate Mind: Structure and Content* (pp. 107-121). Oxford: Oxford University PRess.

Sarnecki, J. (2004). The multimedia mind: An analysis of Prinz on concepts. *Philosophical Psychology, 17*(3), 403-418.

Schouten, M., & Looren de Jong, H. (Eds.) (2007). *The Matter of the Mind: Philosophical Essays on Psychology, Neuroscience and Reduction*. Oxford: Wiley Blackwell.

Shapiro, L. A. (2004). *The Mind Incarnate*. Cambridge, MA: MIT Press.

Skinner, B. F. (1938). *The Behavior of Organisms: An Experimental Analysis*. New York: Appleton-Century.

Solomon, K. O., & Barsalou, L. W. (2001). Representing properties locally. *Cognitive Psychology, 43*, 129-169.

Sperber, D. (1994). The modularity of thought and the epidemiology of representations. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Domain Specificity in Cognition and Culture* (pp. 39-67). Cambridge: Cambridge University Press.

Sperber, D. (2001). In defense of massive modularity. In E. Dupoux (Ed.), *Language, Brain and Cognitive Development: Essays in Honor of Jacques Mehler* (pp. 47-47). Cambridge, MA: MIT Press.

Stich, S., & Nichols, S. (1992). Folk psychology: Simulation or tacit theory. *Mind & Language, 7*(1), 35-71.

Stich, S., & Nichols, S. (1995). Second thoughts on simulation. In M. Davies & T. Stone (Eds.), *Mental Simulation: Philosophical and Psychological Essays* (pp. 87-108). Oxford: Blackwell.

Stich, S., & Warfield, T. A. (Eds.) (1994). *Mental Representation*. Oxford: Blackwell.

Thaler, R. H. (1980). Toward a positive theory of consumer choice. *Journal of Economic Behavior and Organization*, *1*, 39-60.

Thelen, E., & Smith, L. B. (1994). *A Dynamic Systems Approach to the Development of Cognition and Action*. Cambridge, MA: MIT Press.

Titchener, E. B. (1899). Structural and functional psychology. *Philosophical Review*, *8*, 290-299.

Tooby, J., & Cosmides, L. (1992). The psychological foundations of culture. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The Adapted Mind: Evolutionary Psychology and the Generation of Culture* (pp. 19-136). New York: Oxford University Press.

Trout, J. D. (1998). *Measuring the Intentional World: Realism, Naturalism, and Quantitative Methods in the Behavioral Sciences*. New York: Oxford University Press.

Trout, J. D. (1999). Measured realism and statistical inference: An explanation for the fast progress of "hard" psychology. *Philosophy of Science*, *66*, S260-S272.

Vera, A. H., and Simon, H. A. (1993). Situated action: A symbolic interpretation. *Cognitive Science*, *17*, 7-48.

Watson, J. B. (1913). Psychology as the behaviorist views it. *Psychological Review, 20*, 158-177.

Weiskopf, D. (Ms). The plurality of concepts.
Wellman, H. M. (1990). *The Child's Theory of Mind*. Cambridge, MA: MIT Press.
Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review, 9*(4), 625-636.
Wilson, R. (2004). *Boundaries of the Mind: The Individual in the Fragile Sciences*. Cambridge: Cambridge University Press.