

**The Mind's Bermuda Triangle: Philosophy of Emotions and Empirical Science.**

Ronald de Sousa, University of Toronto.

**ABSTRACT**

Philosophy is the name we give to questions for the solution of which there is no methodological consensus. When such consensus arises, a science branches off. Mathematics, physics, biology, and psychology are well-known examples. But since Quine's "blurring of the supposed boundary between speculative metaphysics and natural science", some philosophers of mind have been keen to recover philosophy's kinship with science.

Conceptual analysis of emotion cannot ignore its physiological and behavioural components. But how far do philosophers need to inquire into the actual details of those components' mechanisms? In approaching this question, I will first sketch a few of the ways in which philosophical and psychological models have attempted to deal with inner conflict. I then present two simple objections to the idea that introspective phenomenology can yield reliable access to meanings. A consequence of the lack of transparent access to the meaning of our own mental states is that much traditional philosophical dispute, rather like the search for "explanations" for mysterious ship losses in the Bermuda triangle, is idle fussing over non-facts. Science-free philosophy, like many of our attempts to explain and justify our feelings and behaviour, is largely confabulation. Nevertheless, science sometimes comes merely to confirm or reframe results already established by careful philosophical analysis. In illustration of this thesis, I examine two related puzzles involving emotions: the role of conscious will, and imaginative resistance. Both, I shall argue, may be viewed as adaptive cases of functional encapsulation.

## I. Introduction

My broad concern in this paper is with the fruitfulness of intercourse between science and philosophy. I will approach this by focusing more narrowly, first, on a sampling of approaches adopted by practitioners of emotion science, and, secondly, on two specific sample problems. Of those two, one represents a time-worn philosophical puzzle, while the other is relatively new. The verdict will be mixed. Throughout the history of our subject, philosophers have claimed certainty for propositions regarded as necessarily true, but contradicted with equal aplomb by other philosophers. In this respect, philosophy's record is only slightly better than that of theology. It's time to say *Nostra Culpa*, and to start emulating the more tentative stance of science. On the other hand, we will find that philosophers have often done quite well at anticipating or framing, by dint of careful analysis, the scientific findings most relevant to our conclusions. Scientific findings sometimes come not to establish so much as merely to flesh out a picture already drawn fairly lucidly by philosophers.

Some empirical facts, if established, must have consequences for our philosophical views. If, for example, it could be shown that *all* of our behaviour is robotically determined by universal dispositions triggered by specific circumstances—with or without some room for play generated by randomness—then our notion of moral responsibility would require some reconstruction.

Conceptions of emotions, like other aspects of the philosophy of mind, are particularly liable to be affected by empirical discoveries. Sometimes we may find ourselves driven to trade off definitional axioms against empirical claims. Commonplace observations appear to contradict conventional assumptions and definitions. One example: we think emotions are grounded in articulable intentional states, and yet we ascribe emotions to infants and animals. We could define the latter away by speaking of “proto-emotions”. Or we could amend the original

assumption in the light of the homologies between the physiological and behavioral dispositions of humans and other animals. Similarly, starting with the fact of “recalcitrant emotions”—such as fear of flying—that appear to embody judgments inconsistent with those we rationally endorse, we can give up the doctrine that emotions embody judgments, or give up the axiom that we can't simultaneously endorse contradictory judgments. I shall argue that a third option should be preferred: we can try to construct an empirically adequate model that removes the apparent contradiction. A baseline constraint on a satisfying explanation is that it rest on empirically ascertainable mechanisms.

A promising framework for understanding many of the puzzles generated by the role of emotions in motivation and behaviour, I shall suggest, is to be found in the idea of the “two-track mind”. This is the hypothesis that our behaviour derives from two relatively independent processing systems with different evolutionary origins in the brain. This approach is currently widely endorsed: a 2004 book lists some two dozen versions (Stanovich 2004, 30), and it is being applied to a surprisingly diverse range of problems including mathematical reasoning and knowledge ascription (Dehaene 1997; Nagel 2008). If indeed there are two types of mental processing, the fact that they sometimes yield incompatible verdicts ceases to be surprising.

## II. Blurring the boundaries

Philosophy's traditionally lofty self-image rests on two props. One is the idea that philosophy's job is conceptual analysis; the other is that by virtue of being a native speaker I have privileged access to the concepts I set about to analyze. On these two grounds, we don't have to pay too much attention to non-linguistic facts. I shall call ‘conceptual autonomy’ the presupposition, going back to Hume, that all truth claims can be classed into one of two groups. Those that pertain to “relations of ideas” comprise the *necessary*, the *analytic*, and the *a priori*. Science, by contrast, is in the business of inquiring into “matters of fact”, consisting in *contingent*, *synthetic*, *a posteriori* truths. Curiously, it seems Hume didn't ask in which category

the distinction itself is to be placed, though he urges that whatever is not one or the other we should “commit ... to the flames: for it can contain nothing but sophistry and illusion.” (Hume 1975, §12 pt III). Sure enough, the two groups have not maintained their apparent orderly coherence. In Hume's stark version all and only relations of ideas are analytic, necessary, and knowable a priori. For Kant, some necessities, such as those of arithmetic, are a priori but synthetic. More recently Saul Kripke has argued that some contingent truths might be known a priori, and that some necessary truths are a known posteriori (Kripke 1980). In illustration of the former, he cites the Standard Metre Bar deposited at the Paris Bureau of Weights and Measures. Although it has now lost its status, this was once the object by reference to which the metre was defined. Any other thing was a metre long iff it was *the same length* as the Meter Bar. Contrary to Wittgenstein, who alleged that it made no sense to ascribe a particular length to the Standard Metre Bar (Wittgenstein 1953, §50), Kripke plausibly maintained that, at any particular time under specific conditions, the fact that the Standard Bar is just *that length* remains contingent; it is known a priori and true by definition, however, that *that length* is one metre. A necessary truth known a posteriori is that water is essentially H<sub>2</sub>O. Essences are necessary, but since they must be discovered they are not analytic.

The dogma of privileged access to meaning is neatly captured by Descartes's claim that I may not always know what I see, but what I *seem* to see “cannot be false” even when I am dreaming (Descartes 1986, Med. II, 29). And, Descartes assumes, I must know what I mean when I say that. This is the germ of the program of Husserl's Phenomenology, which aims to identify essences by attending to pure experience, stripped bare by “epoche” or suspension of ontological assumptions (Husserl 1960).

In the light of recent philosophical and scientific work, both conceptual autonomy and privileged access must be abandoned. Since Quine's 1951 “blurring of the supposed boundary between speculative metaphysics and natural science” (Quine 2006, 20), many philosophers of

mind have been keen to recover philosophy's kinship with science. Quine rejected the absolute distinction between synthetic and analytic propositions. Semantics evolve in the service of pragmatic ends. Making adjustments in our understanding of words is sometimes more expedient than changing our beliefs about empirical facts. Particular observations and general principles are tested against one another, and adjustments can go either way. Others have explored the significance of this blurring of boundaries. Ruth Millikan has attacked “meaning rationalism”, the doctrine that we have incorrigible access to our own meanings—specifically meaning identity and difference, univocity and meaningfulness. (Millikan 1993). In the spirit of that work, but omitting much detail, I now sketch two arguments against the program of phenomenology, interpreted as the doctrine that skilled introspection can give reliable access to the character and meaning of one’s own mental states or dispositions.<sup>1</sup>

### III. Two arguments against phenomenology

It is standard procedure in “conceptual analysis” to rely on intuitions about “what we would say”. We might ask: “What would you say to someone who was sincerely convinced that she had seen a round square?” “If someone had committed a crime under the influence of a post-hypnotic suggestion, would you say she was responsible for that crime?” Or perhaps: “What

---

<sup>1</sup> The arguments in the next section are narrowly targeted at phenomenology as a method capable of yielding certainty about mental states and meanings. They are not intended to impugn the usefulness of introspection as an adjunct to cognitive science. It is clear that paying careful attention to the features of experience might disclose constraints on the possible physiological or computational explanations of those experiences, just as those physiological processes themselves limit the possible interpretations of those experiences. There are active debates about the role of phenomenology in the cognitive sciences. See (Gallagher and Varela 2001), and contributions to a special issue of *Phenomenology and Cognitive Science* (notably Roy 2007; Siewert 2007; Thompson 2007; Dennett 2007). Among the disputed aspects of phenomenology are the irreducibility of the first person, and the question of whether phenomenologists expect the analysis of experience to yield knowledge of the underlying mechanisms. I take no position here on those broader issues. (But see de Sousa 1999).

would we say if we were being addressed in English by a living and apparently ordinary tortoise?" This procedure raises many questions, not least of which are whom we should choose to ask, and why we should care about the answer. But the problem that goes to the heart of the phenomenological method is that the answer you get to the hypothetical question "if p, would you say (or do) q?" may not be indicative of what you would actually do or say if p were known to be true. Assenting to 'if p, I would assent to q' does not establish that one would, on learning that p, actually assent to q. This simple point suffices to establish that correct answers to such hypothetical questions about one's mental dispositions are not reliably accessible to consciousness. Our systematic inability to gauge our past and future evaluations is but one aspect of our ignorance of our motives and dispositions, a lacuna in our conscious life which cognitive science has abundantly established. (Gilbert 2006; Wilson 2002).

It might be objected that the limitations just noted concern only our *future* emotions and behaviour. We might still have incorrigible insight into our *present* states, if only these could be adequately delimited. But the second argument disposes of that possibility, showing that we don't even have incorrigible access to the meaning of our current beliefs. The referential content of a belief depends on the meaning of the terms in which I would express it. That in turn depends on factors that are beyond my knowledge or control. This thesis of "externalism"—in one of several senses that have been the topic of philosophical controversy—was first put forward by Hilary Putnam in an influential paper (Putnam 1975) in which he famously argued that "Meanings ain't in the head". The meaning of a speaker's utterance depends in part on causal or referential facts that may be unavailable to that speaker. In particular, the denotation of a term commonly depends on the identity of the thing or stuff to which other speakers first intended to apply it. If so, then I can have no incorrigible knowledge of that referent's nature. To the extent that the referent's original nature partly determines the meaning of what I have "in mind", then, I do not have incorrigible knowledge of my own meaning.

Here is a different way of making the same point. Meaning presupposes some idea of function: the meaning of a term relates to the function that the term serves in our linguistic economy. But we can identify a function only in the context of the larger environment in which that function is effected. Inspection of a pair of scissors in isolation would not reveal their function. We need to see them at work with paper, thread, or dough. Similarly, any meaning function presupposes a context that includes, for each word used, contrastive and distributive sets. The meaning of 'blue' depends in part on all the colour words that might fit in the same contexts, as well as on the sorts of sentential contexts in which 'blue' and its contrasts could figure (Ziff 1964). Not all of that is necessarily available to me right now. I cannot have present experience of the larger context that provides my experience with its full meaning. Indeed, if I did have access to it, understanding the functional role of that context would presuppose a still larger context and so generate a regress. The experience of meaning is not in general capable of revealing the fullness of meaning.<sup>2</sup>

The foregoing arguments are not examples of empirical facts affecting philosophy. Instead they are conceptual arguments for admitting the empirical into philosophy. They are of particular relevance to emotions because emotions are typically *felt*. One might therefore assume that they are best explored from the subjective point of view. But insofar as an emotion's nature involves meanings, the arguments just sketched entail that what emotions feel like cannot give us full access to their nature. Emotions are also characteristically the most obviously embodied of our mental states. Their manifestations are both mental and physical, and theories of emotion over the last century and a half have distributed themselves along the path of a pendulum swinging between the physiological and the intentional poles. The first is represented by William

---

<sup>2</sup> It is worth noting that the same argument refutes rather than supports any type-identity of brain states and meanings. A neural state does not wear its own history on its face. But that history might be crucial to the determination of its function. The function and meaning of a state S might be different, for example, depending on whether S is reached on the way from A to B or from B to A.

James (1884) and recently by Damasio (1999); the other by “cognitivists” such as Robert Solomon (1984; 2007), or Martha Nussbaum (2001) whose view of emotions stresses judgment, intentionality, and even choice. But emotions are also historically determined; and we have just seen that the reason we don't have full access to the meaning of our own mental states is precisely that they are in part historically determined. It follows that emotions are not altogether knowable on the basis of what is available only in the moment's consciousness. Indeed, emotions are shaped by history on at least two levels: that of individual development from zygote through infant to adult, but also on the phylogenetic scale of our evolution into the animals we are. Consequently it is not surprising that emotions are not entirely transparent to consciousness.

#### IV. Emotions and the Two Track Mind

The traditional view of emotions as *disruptive* gives them a central role to play in the experience of inner conflict, which in turn has motivated some of the best-known models of the dynamics or structure of the person in philosophy and science. Plato posited three parts of the soul, arguing that emotion could not be identified with either reason or desire since it sided sometimes with one and sometime with the other (Plato 1997, Bk VII). Freud also favoured a three-part model, but did not assign a clear place for emotions, which span all three of Id, Ego and Superego (Freud 1923). We get yet another three-part model in neuroscientist Paul MacLean's (1975) “triune brain”. MacLean suggested that the structure of our brains reflects a phylogenetic development in which successive layers partly replaced but also partly reduplicated functions established in earlier structures. A primitive “crocodile” brain controls basic autonomic bodily functions and is homologous with that of even distant vertebrate cousins; the limbic system, which we have in common with close mammalian relatives, is responsible for the more instinctual dispositions built into our emotional repertoires; and our uniquely developed neo-cortex takes care of so-called “higher” functions. The limbic system is commonly assimilated to the “emotional brain”, and there is plenty of evidence for the importance of those brain regions



to at least “basic” emotions or “affect programs” (Griffiths 1997). But emotions more loosely defined involve cortical regions as well. Another version of the “divided brain” is the two-hemisphere theory that Gazzaniga (1970) and others inferred from the effects of commissurotomies undertaken to impede the progress of particularly devastating epileptic episodes. Commissurotomy patients appeared in some respects literally to have “two minds”. There is some evidence that left-hemisphere dominance was associated with a propensity to depression, while right-hemisphere dominance in the presence of a left side lesion has been reported to leave a person more cheerful (Taylor 2006). But none of that work gives us reason to attribute a more significant role to one hemisphere in emotions generally.

The lack of clean fit between any of those models and the great variety of what we are wont to call emotions suggests that emotions are too diverse to be pinned to a specific part of the brain. Many of the features we associate with the modus operandi of emotions, however, can be classified as belonging to the First or Intuitive Track in the two-track model mentioned above. The contrast between an ‘intuitive’ and an ‘analytic’ mode of processing is reminiscent of the Freudian distinction between primary and secondary processing (Freud 1911). For Freud, however, primary processing was something we needed to get over very quickly, because it was driven exclusively by the “pleasure principle”, dominated by the search for instant gratification. It was guided by phantasy rather than by any ability to track the outside world. In modern formulations of the Two Track mind, both the intuitive and the analytic types of mental functioning are vital parts of our normal adult life. Intuitive processes are sometimes thought to be modular; they are typically associative, holistic, relatively fast, parallel, automatic, cognitively undemanding, highly contextualized, and organized around stable, “short-leashed” goal structures. Analytic processes, by contrast, are rule-based, controlled, serial, cognitively demanding, relatively slow, decontextualized, and organized around “long-leashed” and

frequently updated goals.<sup>3</sup> The analytic system is typically dependent on explicit linguistic or at least digital representation. This explains why computers are much more easily programmed to perform the latter class of tasks, including ones we find very hard, such as complex analytic calculations. Tasks we find easy—walking, catching a ball, recognizing a face, reading handwriting—have turned out to be hardest to program. This would greatly have surprised Descartes, who thought that animal functions would be easy to replicate in automata whereas anything resembling reason would be impossible. But to us it should not seem surprising, remembering that arithmetic and other analytic systems are recent inventions, whereas the intuitive systems that mostly keep us alive from day to day have evolved over millions of years, albeit often in a “klugey” way (Marcus 2008), and we don't in general know exactly how to reverse engineer them. Notice, however, that while we, like Descartes, continue to regard explicit thinking as essentially human, most of our life skills are governed by the intuitive track. As Paul Churchland has remarked,

“We are able to discriminate and recognise faces, colors, smells, voices, spatial relations, musical compositions, locomotor gaits, the facial expression of emotions, flowing liquids, falling bodies, and a million other things besides, where the exercise of these discriminative skills is once again largely or wholly inarticulable” (Churchland 2007, 90).

## V. Models of emotion

Emotions have been fitted into these and other models in a bewildering variety of ways. Some start from terms familiar to us from folk psychology: desire, belief, will. Others start with what are taken to be simpler component states such as arousal or activation, and valence or

---

<sup>3</sup> See Stanovich (2004); also Strack and Deutsch (2004), whose “Impulsive” and “Reflective” systems correspond roughly to my “Intuitive” and “Analytic”.

degree of pleasantness or unpleasantness (Russell 2005). Robert Plutchik (2001) arranges terms from the repertoire of folk psychology onto a concentric wheel, generating complex emotions as blends of simpler ones, comprising surprise, fear, sorrow, disgust, anticipation, anger, joy and acceptance. The same list, minus ‘anticipation’ and ‘acceptance’, forms the six “basic emotions” that have, according to Ekman and Friesen (1975), a claim to universality of both experience and recognition. Yet another approach, based on a number of different dimensions of appraisal, is represented by Klaus Scherer (2005), who recently offered the following definition of emotion:

an episode of interrelated, synchronized changes in the states of all or most of the five organismic subsystems in response to the evaluation of an external or internal stimulus event as relevant to major concerns of the organism.

Scherer's five systems are: *evaluation of objects and events; system regulation; preparation and direction of action; communication of reaction and behavioral intention; and monitoring of internal state and organism-environment interaction*. In experience, these correspond respectively to the cognitive component; neurophysiological events; motivation; a motor expression component; and subjective feeling. (Scherer 2005, 697–8). For Jaak Panksepp's (2001), there are four major systems controlling emotion: a “seeking system”, corresponding to joy; an “attack system”, and two more, “fear” and “panic,” which common speech doesn't distinguish very clearly, but which differ as threat of harm differs from threat of abandonment in roughly the sense of Bowlby (1969–1980). These cut across Scherer's five “components”: each system comprises a coordinated set of genetically designed hormonal modulators and neural circuits, involving specific regions of the brain, as well as characteristic behavioural and experiential aspects.

It should be clear from this partial sampling that while there is not one single scientific approach to the modeling of emotions, emotions theory has not unequivocally emerged from the methodological chaos that entitles it to be called philosophy. The diversity of these models suggests a need, in Locke's apt phrase, for conceptual "under-labourer[s]... removing some of the rubbish that lies in the way to knowledge." (Locke 1975, 9–10). Nevertheless it seems unpromising to philosophize about emotions without an eye on what those various scientific approaches can teach us. I shall suggest that the complications entailed for our conception of emotions will bring some gain in our potential understanding, but also some risk of loss.

## VI. Free-Will

I move now to the first of my illustrative puzzles: the problem of the will in science and philosophy. The will presents us with the classic case of inner conflict in "weakness of the will" or *akrasia*, which motivated Plato and others to postulate the "divided soul" models mentioned above. But I focus not on conflict but on the very idea of a will.

Sartre (1993) famously said that we are "condemned to be free." If I ask you to wag your finger, you might or might not comply, but you cannot avoid deciding. If you wag your finger, that is a decision; if you refuse, it is one too. And if you simply ignore me, that too is a decision. Anyone observing you can just wait and see what the random processes or causal chains that govern your behaviour will determine you to do; but you yourself cannot.

Nevertheless, Daniel Wegner (2002) has presented a number of ingenious experiments to demonstrate that conscious will is an illusion. Some of his manipulations replicate table turning, in which agents act while entirely convinced of having merely "followed" the table's movements without any causal input whatever. Wegner also devised experiments to generate the opposite illusion, in which subjects are convinced they are doing something which in actual fact is entirely out of their control. This shows that the subjective sense of control is not always veridical. Further evidence for this has come from the now well-known experiments of Benjamin Libet

(Libet, Wright, Feinstein, et al. 1979), which showed that in situations like the simple finger-wagging case above, the consciousness of deciding lags almost a whole second behind the activation of the readiness potential that signals that the machinery of motion has already been triggered, and about 200-300ms after the activation of the brain's motor centres. We are forced to conclude—on pain of giving up the axiom that causes precede their effects—that free-will cannot be assimilated to the causal efficacy of conscious decision. On the contrary, “deciding” is no less an effect of something else than the act itself. In one more illustration of the self-ignorance already noted, Wegner surmised that the consciousness of having willed something is often a confabulation, devised after the fact on the basis of common-sense assumptions about what would justify behaviour of that kind.

Some more news from neuroscience, however, may mitigate this demotion of the role of consciousness in the elaboration of voluntary behaviour. Patrick Haggard successfully replicated Libet's result, but suggested that while the overall planning of an act undoubtedly precedes consciousness, we can ascribe to the latter a monitoring and predictive role in the detail of the plan's execution: “conscious intentions are at least partly preconstructions, rather than mere reconstructions.” (Haggard 2005, 293). This idea is supported in (Shariff and Peterson 2008), who argue that we can keep as much free-will as anyone ought to want providing we recognize two different mechanisms in the brain, one of which executes a plan elaborated in the longer term by the other. The executive phase, which does not involve consciousness, is a First Track process, even when it has been set up by First Track deliberation. (See also Clark 2007). This fits the phenomenology of action nicely. If you are a skier, Shariff and Peterson point out, you look not at the tip of your skis but several yards beyond you. What your skis are doing now was determined by your look ahead a few fractions of a second ago. Indeed, in new work by John-Dylan Haynes and his colleagues at the Max Planck Institute, using fMRI, the brain has been found to have embarked on an action as much as 10 seconds before the moment of conscious

“decision”. (Soon, Brass, Heinze, et al. 2008). In effect, then, what we interpret as the present is really the future: subjective experience misconstrues what is actually an anticipation and reads it as present consciousness. The awareness of a stimulus is attributed to the present, but it actually targets the future and is referred back. Conscious time is out of step with real time.

The experiments just cited are impressive, and it is hard to deny that they must have philosophical consequences. But it isn't easy to say what these are. To begin with, they don't suffice to establish that “there is no free-will”: that depends on just what that term is taken to imply. It need not, in particular, include the requirement that the moment of conscious decision must be the determining causal factor in ordinary action. That cannot be the case, since it takes place after the whole mechanism of action has been set in motion. But we could know that by just thinking about it. A classic argument runs as follows. Libertarians require that a free will must be either independent or fully conscious of any determining factor. If I claim sole responsibility for an action, I should be entitled to affirm that I originated or consciously took account of all the factors that have determined that decision. If my origination of the act was unrelated to any previously existing wants, inclinations, or prior intentions, then it cannot count as mine. If, on the other hand, my act did have immediate causal antecedents in my states of mind and character, then those states in turn may have been determined by others of which I am not and could not possibly be aware. In short, the libertarian requirement is incoherent.

In light of that argument, we didn't need the science to tell us that conscious will or decision cannot be the uncaused origin of action. Once again, the brain data may be little more than interesting elaborations of a point already worked out correctly by philosophy—in this case by Hume, who pointed out that chance, not freedom, is the opposite of necessity (Hume 1975, §VIII Pt I). Chance is not what libertarians want. They think there must be room here for absolute origination by the self, a pure act of uncaused creation that is neither determined nor random. Hume couldn't understand what this amounted to; I can't understand what it means

either. Libertarians think they can make sense of it (Wiggins 1987; van Inwagen 1983). To avoid mere invective or accusations of irrationality, it is perhaps best to concede that both sides may be manifesting differences of philosophical temperament which will not be bridged by argument alone. This attests to the reality of intellectual temperament, which remains at the root of many differences of opinion even if one has abandoned the theological certainty I deplored above.

The mutual incomprehension of compatibilist and libertarian also affords an entry point into my last illustrative case, by providing a little noticed example of an *inability to imagine* how someone could hold a given position. Several related phenomena have been discussed under the heading of “imaginative resistance”. The ones I shall focus on are very different; but it is well to keep in mind the parochial case of mutually unimaginable philosophical positions.

## VII. Imaginative Resistance

Imagination is a tool for the exploration of the possible. But what is possible and what is imaginable are not the same. Imaginary numbers, for example, despite their name, are not imaginable. Neither are higher dimensionalities. Yet both are well conceived mathematical realities. The converse is also false: to some extent, we can imagine some impossible things. When you are telling a story, the imagination is not limited to what is possible physically, or chemically, or even logically. Most impossibilities, if not all, can at least be “conceived”, if only well enough for their impossibility to be clearly grasped. Imagination connotes *imaging*, though it isn't always understood to require it; hence the importance of sensory content when we imagine what exactly imagining might be. And yet there are limits to what can be imagined, as the case of philosophical incomprehension has just shown. To raise the question of the possibility of imagination is to presuppose that imagination can fail: where the limits of imagination are depends in part on what counts as *successful* imagination (Morton 2006). If I am asked to imagine an object specified in a self-contradictory formula, for example, I might claim to comply by listing successive observations of the inconsistent properties. Yes, I can imagine a

round square: it's all smooth and uniform, and then I can feel the sharpness of the four corners. I might even be able to imagine seeing that a box is empty and that it contains a carved figurine (Priest 1997). Escher's prints represent impossible spatial relations and so can be thought to imagine them. And in Roy Sorensen's amusing challenge to provide a picture of an impossible object, he is at pains to preclude bogus entries such as this dot, ‘.’ as a picture of an impossible object viewed from very far away (Sorensen 2002).

The general question of the limits of imagination is therefore complex, and too hard for me to sort out. My interest here is much narrower, and is motivated by my general concern with the contributions of science to the philosophy of emotion. The puzzle I want to focus on poses not so much problems of definition or logic, but a specific form of the problem of will—it is, roughly speaking, a problem about the constraints on wanting and emoting at will.

The puzzle I am concerned with is commonly said to have originated in a much cited passage where Hume describes our response to writings about other ages or places:

There needs but a certain turn of thought or imagination to make us enter into all the opinions, which then prevailed, and relish the sentiments or conclusions derived from them. But a very violent effort is requisite to change our judgment of manners, and excite sentiments of approbation or blame, love or hatred, different from those to which the mind from long custom has been familiarized. And where a man is confident of the rectitude of that moral standard, by which he judges, he is justly jealous of it, and will not pervert the sentiments of his heart for a moment, in complaisance to any writer whatsoever. (Hume 1965, §33).

The thought is that it is easy to imagine strange people doing extremely strange things, but that “a very violent effort” is needed to imagine oneself *approving* of what they do. What is not clear from Hume's remarks, however, is whether his allusion to



effort implies that we may sometimes be literally *unable* to imagine endorsing certain judgments, or rather that we *ought* not do so: “I cannot, nor is it proper I should, enter into such sentiments”, he writes in the preceding paragraph. If his claim is that we *should* not imagine certain things, he may be endorsing Plato’s reason for banishing art, namely that art enlists the imagination in the promotion of bad behaviour. The tradition that rehearses this argument is alive and well even now (Inderscience 2008). Alluring though it is, I shall make every effort to evade that debate. Since ‘should’ is often held to imply ‘can’, some have taken the view that imaginative resistance is exclusively a matter of unwillingness rather than inability. (Gendler 2000, 2006). In that guise, the puzzle raises questions about what we *want* to imagine; but it also invites us to ask whether there are constraints on *what we can want* to imagine. That aspect of it connects with a classic problem about the compatibilist conception of free-will: whether some of my wants are deeper or more deeply “mine” than others, and how to tell which those are. And that, in turn, evokes a certain model of the self—as a sort of onion, perhaps, with endless concentric layers. Purely on the basis of conceptual analysis, then, even that narrow form of the puzzle leads us far afield.

Although Hume's discussion invoked specifically moral attitudes, the difference between moral and non-moral attitudes will be marginal to my concerns. I shall also ignore the question of the aesthetic value of immoral art or pleasure, except to acknowledge that the purest aesthete may find it difficult to circumscribe the aesthetic domain. Even the most hedonistic gourmet's pleasure in eating chocolate may be spoiled by the thought of the conditions under which cocoa beans are harvested. But that sort of interference may have more in common with distraction than with imaginative resistance—best compared to the difficulty of concentrating on your reading in a noisy environment. The difficulty that attends the sort of cases I am looking for is unlike this, and also unlike the difficulty I might find in imagining four-dimensional objects.

What I am unable to do is not to frame the *content* of an imagined situation, but to *respond* in certain ways to certain imagined prospects. That crucial distinction is nicely elucidated by Peter Goldie (2003, 57), who gives the following illustration: “the disgust and horror I now feel at my [imagined future] self, old, decrepit and senile, is a *response* to what I imagine; it is not part of the *content* of what I imagine.”<sup>4</sup>

Two cases will most clearly make my point: *sexual arousal* and *amusement*. I know of, and to some extent can visualize (with or without the assistance of the *News of the World*) sexual practices that others enjoy, but which entirely fail to arouse me. Similarly, the same joke or event can move one person to laughter which, although I can see the point, leaves me quite cold.<sup>5</sup>

Here is a joke, related to Jenny Diski by a South African acquaintance:

*‘An Englishman, a Thai and an African were all together at Oxbridge. After some years the Englishman goes to visit the Thai who is hugely rich. How come? asks the Englishman. See that road? I own 10 per cent of it, the Thai tells him. The Englishman goes to visit the African, who is also hugely rich. How come? See that road? says the African. What road? the Englishman asks.’* Moira waited for me to burst out laughing, but it was a minute or two before I could make anything at all of this story. Besides, what were the overseas students doing in Oxbridge in the first place if they weren't rich already? (Diski 2008).

---

<sup>4</sup> See also Moran (1994). Goldie also rightly notes, borrowing an important distinction from Richard Wollheim (1984) that the same event can be either centrally or acentrally imagined, even where I cast myself as a protagonist in the imagined event. Depending on whether I am imagining centrally or acentrally, I might feel fear or compassion. (ibid).

<sup>5</sup> The target of the amusement doesn't have to be a joke. Here is a case where the target of laughter is an actual event. It is reported that at a performance by child acrobats attended by Chairman Mao, one young boy fell and was badly hurt. The crowd cried out in horror—except for Mao, who laughed. (Mirsky 1994, quoting Li Zhisui 1994).

Given a minute or two, Diski sees the point, and there is no obstacle to imagining the scene, set up as in any classic joke. But what she doesn't do is *find it funny*. Finding it funny *in that context* would require sharing the teller's perceived attitudes, including the assumption that making money out of failing to build a road is typical of Africans, but perhaps contempt and resentment towards Africans in general. Suspension of disbelief is easy; but it is powerless to induce the attitude that is required to find something funny.

The qualifier “in that context” is important; Diski might have laughed if she had heard it from an African whose attitudes were subtly different. And one sometimes laughs at a joke for its sheer cleverness, despite one's disapproval of the underlying attitude. One may laugh at a joke one disapproves of, and one can even approve of laughing at jokes one disapproves of (Jacobson 1997). But the crucial point here is, using Goldie's useful distinction again, that our power to imagine a given *content* is virtually unlimited regardless of context, whereas our emotional *response* is not equally under our power.

What is true of amusement is also true of sexual arousal. Most people can think of *something* they find disgusting that some others—and perhaps even themselves, in other moods or other contexts—find arousing or actually do for pleasure. For those people, or in those moods, it might seem impossible to imagine *enjoying* the joke or the sexual practice. But what is really going on in these cases is not so much an inability to imagine, but to experience a certain *emotional response* (or an absence of emotional response) to something imagined. A further complication is that one's imagination of the sexual scene (whether central or acentral) can include arousal in its *content*. And – as Peter Goldie has pointed out to me – while imagining one's own arousal may itself be arousing, a failure of response in the present might make it more difficult to imagine the arousal response as part of the content.

## VIII Lessons from psychopathy

As in all cases where likings or tastes are in question, we can wonder about the separability of cognitive and emotional responses. Psychopaths are widely reputed to be capable of reasoning impeccably and arriving at correct moral judgments, while altogether failing to be moved to act or care. We can describe the psychopath in either of two ways. One says that his avowals are merely insincere. The other, while deploring his failure to be moved, grants that he may be competent and sincere, since he can articulate the implications of the statement as well as anyone else. There is no cognitive content he is intentionally withholding. Each description appears to beg the question against the other. The second ignores the reasonable requirement of a condition of sincerity, where sincerity implies caring. But that claim itself assumes without proof that evaluative emotions supervene on the character of their cognitive base. It is at least logically possible that two people might have qualitatively identical sensations and cognitions in all respects except for the fact that one responds with liking, and the other with dislike or indifference.

Whether and to what extent natural facts under-determine value judgments remains a live if somewhat metaphysical dispute. But recent findings about the brains of normals and psychopaths, as well as about normal brain activity under different circumstances, argue for a genuine gap between fact and value, or at least between cognition and caring. There are also differences in moral opinions between individuals belonging to different cultures and different times. At least it seems that way: the anthropological evidence is never wholly conclusive, however, since no sufficiently exhaustive descriptions could ever be given to show that the “situational meanings” of a belief or attitude were the same for two individuals (Moody-Adams 2002). No observation could conclusively refute the hypothesis that, for values as for tastes, approval and disapproval supervene uniquely on qualitative experiences: *If you don't like (or approve of) it then it doesn't taste (or look or seem) the same to you as it does to me.*

Even the evidence about psychopaths' brains remains inconclusive. At first sight, true psychopaths appear to be capable of saying all the right things; they simply fail to have the physiological responses associated, for “normal” subjects, with fear and empathy. They fail to be *moved* by their apparently evaluative endorsements (Blair, Mitchell and Blair 2005). Here again, there seems to be room for a battle of intuitions about the limits of philosophical imagination. (Dennett 1991). A zombie is a creature that gets full marks on the Turing test. A zombie's utterances are perfectly convincing as human utterances, even when discussing the most subtle shades of experience, but the zombie has no consciousness and experiences nothing. The psychopath seems to qualify as a *moral zombie*: able to manifest all the verbal signs of moral consciousness, without any of the expected emotional responses. He says things that would normally indicate conscious moral experience, or *conscience*, while actually having none at all.<sup>6</sup>

In fact, however, the psychology and neurology of real psychopaths spare us the stalemate over the conceivability of moral zombies. True psychopaths do not actually pass the moral Turing test. They betray themselves in a subtle and curious way, in that their moral judgments reveal their deficit in the very act of trying to conceal it. Most subjects—from fairly early childhood—distinguish between prohibitions that derive from social convention such as etiquette and those that derive from the avoidance of harm and suffering (Turiel 2002). But when psychopaths are asked to classify different prohibitions, they tend to rank them all, even those that most clearly derive from social convention, with absolute prohibitions of harm. They do so, apparently, precisely because they are attempting to conceal their inability to feel the force of prohibitions based on harm. (Prinz 2007, 44). Neuroscience has begun to identify the brain peculiarities that underlie those differences. Physiological measures show that the psychopath

-----  
<sup>6</sup> Jesse Prinz makes this point by adapting Frank Jackson's famous thought experiment about colour scientist Mary, who has never been exposed to an actual colour experience. Prinz's variant is an ethical expert who has never previously had the emotional experience of responding in a way that is appropriate to the judgments she expertly makes. (Prinz 2007, 38ff).

lacks distress responses in the face of others' suffering or his own imminent pain, correlated with a failure of the mechanisms of negative learning. (Blair, Mitchell and Blair 2005)

Neither can we refute, however, the theoretical presumption that the relation between the qualitative experience and the emotional response is contingent—that judgments of value do not supervene on perceptions of fact. No necessity, logical, physical or psychological, precludes different emotional responses from being aroused by the very same qualitative experience. Although the psychopath's brain is differently wired, his experiences might be qualitatively identical from others' in all respects *except* for emotional response and motivation.

We seem to be up against a stalemate. Can we do better?

Recall that in discussions of the puzzle of imaginative resistance some see genuine *inability* to imagine, while others see merely *unwillingness* to do so. I have tried to isolate two robust types of cases involving genuine inability or extreme difficulty. I venture to surmise that the temptation to view these cases as involving unwillingness springs from the fact that *we cannot want at will*. Given the independent reasons I have sketched for regarding the “will” as a concept too murky to shed light on anything at all, I propose, following hints offered by both Matravers (2003) and Currie (2002), that many of the phenomena under discussion can be explained in terms of the adaptive role of imagination in planning.

Our faculty to imagine counterfactual situations enables us to explore alternative scenarios and possible outcomes, in a process akin to computer “simulation” (Oatley 1999). This, in the happy phrase often attributed to Karl Popper, “allows our hypotheses to die in our stead.” (Dennett 1996). Such exploration has the function of providing information about a course of action's likely outcome, but also, crucially, about the significance of that outcome for the agent. If I were to abandon my attitudes when I explore the space of possibilities, the simulation would not serve its purpose. Hence we should not be surprised that our attitudes are more difficult to modify at will than the purely representational aspects of imagination.

The difficulty in question may, however, be a matter of degree. We've seen that our attitudes are not absolutely inflexible: they can vary according to context. We should be able to learn something from the nature of those variations. Simple observation, as well as some indirect evidence from brain studies, suggests that our attitudes are most resistant to change in the light of counterfactual imaginings when they are most likely to require us to *do* something. In fiction, our attitudes are safe from commitment: we can't be expected to do anything about it, and so we can allow ourselves a broader range of sympathies than we can in active life. But the same applies to cases of actual belief about remote events. Plato's thought experiment about the ring of Gyges would be ineffectual if the average person could not fantasize without too much guilt about enjoying the powers it conferred, but we might not feel comfortable doing the same with a real tyrant closer in time and space. Fiction can indeed serve to broaden our sympathies as well as our imagination; but the more extreme the divergence between our own and the imagined attitudes, the more secure we need to be in the thought that "there is [no] risk of being drawn into action" (Goldie 2003, 68). One consequence of this, as Goldie noted, is that we may sometimes care less about real people one personally knows than about fictional ones. A related point, yet sufficiently different to complicate the matter further, is made by the character of Pegeen in the climactic scene of Synge's *Playboy of the Western World*. Pegeen, who thought she admired Christy for murdering his father, changes her mind entirely when he actually splits his father's skull. "And what is it you'll say to me," he cries, "and I after doing it this time in the face of all." "I'll say," she replies, "a strange man is a marvel, with his mighty talk; but what's a squabble in your backyard, and the blow of a loy, have taught me that there's a great gap between a gallous story and a dirty deed." (Synge 1911, III/223–224) Her answer underlines the essential unreliability, noted above, of assent to conditionals as an indicator of conditional assent. It also shows that the difference doesn't line up neatly with the distinction between fiction and practical imagination. In the case of real life, it is more important to protect our enduring attitudes through the thought experiment. Fiction, by contrast, can explore outlandish situations that might entail radical

changes in our own values. Thus fiction can indeed take us to mental places closed to ordinary practical deliberation. This important fact tends to be forgotten by crusaders against violent and sexually aggressive games or movies. Many who enjoy fantasy games as fiction would be repelled by the prospect of acting them out in real life.

Functional MRI explorations of the brain lend some credence to this view of what causes our resistance to shifts of attitudes. When subjects consider different versions of the “trolley problem”, their response is driven more purely by emotion in proportion as their own involvement in the envisaged scenario gets more personal. Most people say they would flip the switch that would divert a trolley from a track where it will kill five people to one where only one will be killed; but the same subjects are mostly reluctant to effect the same result by physically pushing a single fat man onto the track. With few exceptions, they reject the consequentialist solution—save five lives at the cost of one—when their personal involvement is more immediate. Different parts of the brain are active when one is making the utilitarian calculation, than when responding to the “deontological” prohibition against causing harm (Greene, Sommerville, Nystrom, et al. 2001). Antonio Damasio and his colleagues confirmed this when they compared subjects suffering from a lesion in the Ventr Medial Prefrontal Cortex (VMPC) with controls with intact brains. In those with damage to the VMPC, an area crucial to emotion processing, the aversive character of personal intervention failed to overcome the utilitarian calculation of overall benefit. A brain lesion in the area responsible for controlling emotional response made subjects more “rational”, that is, more consistently consequentialist. (Koenigs, Young, Adolphs, et al. 2007). Ironically, in light of Kant's disdain for emotions as motivators of authentically moral acts, this suggests that only consequentialists are moved by reason, while strict Kantians are responding purely to emotion (Haidt and Bjorklund 2008).



## IX. Conclusion

Jerry Fodor (1983) argued that certain functions of the mind were “modular”, or encapsulated, that is, relatively immune to correction from other channels. Thus we see the stick as bent at the surface of the water even when we know it isn't, and we do not make visual illusions go away by ascertaining that we are not seeing them as they are. Perceptions are modular in this sense, but Fodor insisted that general cognition is not. In terms of the Two-Track model, this could be rephrased to say that First Track processes are modular, whereas Second Track ones are not. Efficient and speedy responses are produced without elaborate deliberation. Modularity can be plausibly explained in terms of a selective advantage conferred to encapsulated First Track processes. Some First Track functions belong to general cognition, however, although they work fast and automatically. In some slightly revised sense, cognition is also to some extent modular (Carruthers 2003). The results of brain observation in the two cases I have discussed—the will and imaginative resistance—support the idea that a certain lack of conscious control may be adaptive. Once the preparatory work of the deliberative function is done, the will must be kept from interfering with the efficient functioning of the executive function. In the course of imaginative simulation, we need to protect our value commitments in counterfactual situations in proportion to the extent that we are personally committed to acting.

Natural selection explains why certain functions are encapsulated, and psychology tells us something about how we might nevertheless modify our responses in light of changing circumstances. The efficiency of our First-Track functions would be at risk if they were too easily influenced by external information; and the phenomena I have discussed strengthen the practical relevance of imaginative simulations by anchoring our evaluations of hypothetical outcomes in present values. To some extent, these explanations are still couched in the vocabulary of folk psychology. But neuroscience demonstrates the mutual independence of mechanisms we thought of as forming a single function. Thus deliberation and execution seem to

be united in the Will, but they turn out to be distinct and empirically separable. Similarly imagination and emotional response seem to function as one in the exploration of counterfactual futures, but they draw apart, sometimes usefully so, when the link between imagination and potential action is loosened.

But we must not forget the limits of our intuitive self-understanding, particularly concerning our own motivations and our own emotions (Wilson 2002). We are subject to massive self-deception and confabulation (Hirstein 2005). The rationale I have postulated for keeping values constant through counterfactual deliberation actually works rather poorly: we tend to err in both our predictions and our memories of our own responses (Gilbert 2006). The way out of this predicament is not by way of first person introspection, but on the contrary by treating ourselves as objects, not as subjects. Introspection needs to be supplemented, if perhaps not replaced, with “heterophenomenology”. (Dennett 2003). In doing that, however, we diminish the hold on our conceptual scheme of the traditional categories of emotion. If our emotions are modular, they don't necessarily function in the ways we our existing emotional concepts lead us to expect. When we unravel the causal mechanisms that make it possible to act, we may lose sight of our notions of free-will, of motivation, and of emotions, not because we are puppets of determinism, but because the mechanisms involved are just too *complicated*. Like ships in the Bermuda triangle, our emotions might seem to disappear because they were never really there.

## References

- Blair, J., D. Mitchell, and K. Blair. 2005. *The psychopath: Emotion and the brain*. Oxford: Blackwell.
- Bowlby, J. 1969-80. *Attachment and loss*. New York: Basic Books.
- Carruthers, P. 2003. The mind is a system of modules shaped by natural selection. In *Contemporary debates in the philosophy of science*, ed. C. Hitchcock. Oxford: Blackwell.
- Churchland, P. M. 2007. What happens to reliabilism when it is liberated from the propositional attitudes? In *Neurophilosophy at work*, 88-112. Cambridge: Cambridge University Press.
- Clark, A. 2007. What reaching teaches: Consciousness, control, and the inner zombie. *British Journal for the Philosophy of Science* 58:563-94.
- Currie, G. 2002. Desire in imagination. In *Conceivability and Possibility*, ed. T. Gendler and J. Hawthorne, 201-21. Oxford: Oxford University Press.
- Damasio, A. R. 1999. *The feeling of what happens: Body and emotion in the making of consciousness*. New York, San Diego, London: Harcourt Brace.
- Dehaene, S. 1997. *The number sense*. Oxford: Oxford University Press.
- de Sousa, R. 1999. Twelve varieties of subjectivity. In *Proceedings of the Sixth International Colloquium on Cognitive Science, ICCS.*, ed. J. Lazarrabal.
- Dennett, D. C. 1991. *Consciousness explained*. Boston, Toronto, London: Little, Brown.
- , 1996. *Kinds of Minds: Towards an Understanding of Consciousness*. New York: Basic Books.
- , 2003. Who's on first? heterophenomenology explained. *Journal of Consciousness Studies* 10(9/10), October:19-30.
- , 2007. Heterophenomenology reconsidered. *Phenomenology and the Cognitive Sciences* 6(3):247-70.
- Descartes, R. 1986. *Meditations of First Philosophy, with selections from the Objections and Replies*. Trans. John Cottingham, introd. by Bernard Williams. Cambridge, UK: Cambridge University Press.
- Diski, J. 2008. Diary: On not liking South Africa. *London Review of Books* 30(13), 2003 July:41.
- Ekman, P., and W. Friesen. 1975. *Unmasking the face: A guide to recognizing emotions from facial expressions*. Englewood Cliffs, NJ: Prentice-Hall.
- Fodor, J. 1983. *The modularity of mind*. Cambridge, MA: MIT Press, A Bradford book.

- Freud, S. 1911. Formulations on the two principles of mental functioning. In *Standard edition of the complete psychological works*, vol. 12, ed. J. Strachey, 213-26. London: Hogarth Press.
- . 1923. The Ego and the Id. In *Standard edition of the complete psychological works*, vol. 19-68, ed. J. Strachey, 3-68. London: Hogarth Press.
- Gallagher, S., and F. Varela. 2001. Redrawing the map and resetting the time: Phenomenology and the cognitive sciences. In *The reach of reflection: Issues for phenomenology's second century*, ed. S. Crowell, L. Embre, and S. J. Julian. Center for Advanced Research in Phenomenology, Intrinsic. [Www.electronpress.com](http://www.electronpress.com).
- Gazzaniga, M. 1970. *The bisected brain*. New York: Appleton Century Crofts.
- Gendler, T. S. 2000. The puzzle of imaginative resistance. *Journal of Philosophy* 97(2):55-81.
- . 2006. Imaginative resistance revisited. In *The architecture of the imagination: New essays on pretence, possibility, and fiction*, ed. S. Nichols, 149-73. Oxford: Oxford University Press.
- Gilbert, D. T. 2006. *Stumbling on happiness*:. New York: Knopf.
- Goldie, P. 2003. Narrative, emotion, and perspective, in *Imagination, Philosophy and the arts*, ed. M. Kierans and D. Lopes, 54-68. London: Routledge.
- Goldie, P. 2005. Imagination and the distorting power of emotions. *Journal of Consciousness Studies* 12(8-10):130-42.
- Greene, J. D., E. B. Sommerville, L. A. Nystrom, Darley. John M., and J. D. Cohen. 2001. An fMRI investigation of emotional engagement in moral judgment. *Science* 293 (2001, 14 September):2105-8.
- Griffiths, P. E. 1997. *What emotions really are: The problem of psychological categories*. Science and its conceptual foundations. Chicago: University of Chicago Press.
- Haggard, P. 2005. Conscious intention and motor cognition. *Trends in Cognitive Sciences* 9(6):290-95.
- Haidt, J., and F. Bjorklund. 2008. Social Intuitionists Answer Six Questions about Moral Psychology. In *Moral Psychology, vol. 2*, W. Sinnott-Armstrong, 181-217. Cambridge, MA: MIT.
- Hirstein, W. 2005. *Brain fiction: Self-deception and the riddle of confabulation*. Cambridge, MA: MIT.
- Hume, D. 1975. *Enquiries: Concerning human understanding and concerning the principles of morals*. 3d ed. Ed & introd by L. A. Selby-Bigge. Revised by & notes by P. H. Nidditch. Oxford: Oxford University Press, Clarendon.
- . 1965. *Of the standard of taste and other essays*. Ed & introd by J. W. Lenz. The Library of Liberal Arts. Indianapolis: Bobbs-Merrill.

- Husserl, E. 1960. *Cartesian meditations: An introduction to phenomenology*. Trans. D. Cairns. The Hague: Martinus Nijhoff.
- Inderscience. 2008. Could videogames reduce rather than increase violence. *Science Daily*, 15 May. Accessed 27 July 2008.
- Jacobson, D. 1997. In praise of immoral art. *Philosophical Topics* 25(1):155-200.
- James, W. 1884. What is an emotion? *Mind* 9:188-205.
- Koenigs, M., et al. 2007. Damage to the prefrontal cortex increases utilitarian moral judgements. *Nature* 05631:1-4.
- Kripke, S. A. 1980. *Naming and necessity*. Cambridge, Massachusetts: Harvard University Press.
- Li Zhisui. 1994. *The private life of Chairman Mao: The memoirs of Mao's personal physician*. Tai Hung-chao and A. F. Thurston. New York: Random House.
- Libet, B., E. Wright, Jr., B. Feinstein, and D. Pearl. 1979. Subjective referral of the timing for a conscious experience: A functional role for the somatosensory specific projection system in man. *Brain* 102:193-224.
- Locke, J. 1975. *An essay concerning human understanding*. Ed. P. Nidditch. Oxford: Oxford University Press, Clarendon.
- MacLean, P. D. 1975. Sensory and perceptive factors in emotional functions of the triune brain. In *Emotions: Their parameters and measurement*, ed. L. Levi. New York: Raven Press.
- Marcus, G. 2008. *Kluge: The haphazard construction of the human mind*. New York: Houghton Mifflin.
- Matravers, D. 2003. Fictional assent and the (so-called) "puzzle of imaginative resistance", p. Imagination, and the arts, 91-106. London: Routledge.
- Millikan, R. G. 1993. White Queen psychology; or, the last myth of the Given, White Queen psychology and other essays for Alice, 279-363.
- Mirsky, J. 1994. Unmasking the monster: The memoirs of Mao's personal physician. Review of The Private Life of Chairman Mao. Li Zhisui. *New York Review of Books* 41(19), 1919 November:Online: <http://www.nybooks.com/articles/2080>. Retrieved 25 July 2008.
- Moody-Adams, M. 2002. *Fieldwork in familiar places: Morality, culture, and philosophy*. Cambridge, MA: Harvard University Press.
- Moran, R. 1994. The expression of feeling in imagination. *Philosophical Review* 103(1):75-106.
- Nagel, J. 2008. Knowledge ascriptions and the psychological consequences of thinking about error. (unpublished) Toronto.
- Nussbaum, M. 2001. *Upheavals of thought: A theory of the emotions*. Gifford Lectures. Cambridge: Cambridge University Press.

- Oatley, K. 1999. Why fiction may be twice as true as fact: Fiction as cognitive and emotional simulation. *Review of General Psychology* 3(2):101-17.
- Panksepp, J. 2001. The neuro-evolutionary cusp between emotions and cognitions implications for understanding consciousness and the emergence of a unified mind science. *Evolution and Cognition* 7(2):141-63.
- Plato. 1997. Republic. In *Complete Works*, ed. J. M. Cooper and D. Hutchinson. Indianapolis: Hackett.
- Plutchik, R. 2001. The nature of emotions. *American Scientist* 89(4):344-50.
- Priest, G. 1997. "Sylvan's box: A short story and ten morals. *Notre Dame Journal of Formal Logic* 38(4):573-82.
- Prinz, J. 2007. *The emotional construction of morals*. Oxford; New York.: Oxford University Press.
- Putnam, H. 1975. The meaning of meaning. In *Mind, language and reality*, vol. 2. Philosophical papers, 215-71. Cambridge: Cambridge University Press.
- Quine, W. V. O. 2006. Two dogmas of empiricism. *Philosophical Review* 60:20-43 (1951). Page ref. is to *From a logical point of view*. 2<sup>nd</sup> revised ed., Harvard U. Press.
- Roy, J.-M. 2007. Heterophenomenology and phenomenological skepticism. *Phenomenology and the Cognitive Sciences* 6(3):1-20. Siewert, C. 2007. In favor of (plain) phenomenology. *Phenomenology and the Cognitive Sciences* 6:201-20.
- Russell, J. A. 2005. The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology* 17:715-34.
- Scherer, K. R. 2005. What are emotions? And how can they be measured? *Social Science Information* 44(4):695-729.
- Shariff, A. F., and J. B. Peterson. 2008. Anticipatory consciousness, Libet's veto and a close-enough theory of free will. In *Consciousness & Emotion: Agency, Conscious Choice, and Selective Perception.*, ed. R. D. Ellis and N. Newton, 197-214. Amsterdam: John Benjamins.
- Solomon, R. C. 1984. *The passions: The myth and nature of human emotions*. New York: Doubleday.
- Solomon, R. C. 2007. *True to our feelings: What our emotions are really telling us*. Oxford; New York: Oxford University Press.
- Soon, C. S., M. Brass, H.-J. Heinze, and J.-D. Haynes. 2008. Unconscious determinants of free decisions in the human brain. *Nature Neuroscience* 11 (13 April):543 - 45. Doi:10.1038/nn.2112.
- Sorensen, R. A. 2002. The art of the impossible. In *Conceivability and Possibility*, ed. T. Gendler and J. Hawthorne, 337-68. Oxford: Oxford University Press.
- Stanovich, K. E. 2004. *The Robot's rebellion: Finding meaning in the age of Darwin*. Chicago: University of Chicago Press.

- Strack, F., and R. Deutsch. 2004. Reflective and Impulsive Determinants of Social Behavior. *Personality and Social Psychology Review* 8(3):220-27.
- Synge, J. M. 1911. *The Playboy of the Western World*. Boston: J. W. Luce. Bartleby.com, 2000.  
www.bartleby.com/1010/. Accessed 2008/07/24.
- Taylor, J. B. 2006. *My stroke of insight*. New York: Penguin.
- Turiel, E. 2002. *Culture and morality*. Cambridge: Cambridge University Press.
- van Inwagen, P. 1983. *An essay on free will*. Oxford: Oxford University Press.
- Wegner, D. M. 2002. *The illusion of conscious will*. Cambridge MA: MIT Press.
- Wiggins, D. R. 1987. Towards a reasonable libertarianism. In *Needs, values, truth*, 269-302. Oxford: Basil Blackwell.
- Wilson, T. D. 2002. *Strangers to ourselves: Discovering the adaptive unconscious*. Cambridge, MA; London: Harvard University Press, Belnap.
- Wittgenstein, L. 1953. *Philosophical investigations*. Trans. G. E. M. Anscombe. New York: Macmillan.
- Wollheim, R. 1984. *The Thread of Life*. Cambridge: Cambridge University Press.
- Ziff, P. 1964. *Semantic Analysis*. Ithaca: Cornell University Press.