

# Neurosentimentalism and Moral Agency

PHILIP GERRANS

*The University of Adelaide*  
*philip.gerrans@adelaide.edu.au*

JEANETTE KENNETT

*Macquarie University*  
*jeanette.kennett@scmp.mq.edu.au*

Metaethics has recently been confronted by evidence from cognitive neuroscience that tacit emotional processes play an essential causal role in moral judgement. Most neuroscientists, and some metaethicists, take this evidence to vindicate a version of metaethical sentimentalism. In this paper we argue that the 'dual process' model of cognition that frames the discussion within and without philosophy does not do justice to an important constraint on any theory of deliberation and judgement. Namely, decision-making is the exercise of a capacity for agency. Agency, in turn, requires a capacity to conceive of oneself as temporally extended: to inhabit, in both memory and imagination, an autobiographical past and future. To plan, to commit to plans, and to act in accordance with previous plans requires a diachronic self, able to transcend the present moment. While this fact about agency is central to much of moral philosophy (e.g. in discussions of autonomy and moral responsibility) it is opaque to the dual process framework and those meta-ethical accounts which situate themselves within this model of cognition. We show how this is the case and argue for an empirically adequate account of moral judgement which gives sufficient role to memory and imagination as cognitive prerequisites of agency. We reconsider the empirical evidence, provide an alternative, agentic, interpretation of key findings, and evaluate the consequences for metaethics.

## Introduction: neurosentimentalism

Many philosophers, and most neuroscientists, take recent work in cognitive science on the neural basis of social and moral decision making to vindicate some version of sentimentalism in metaethics. The simplest version of this view is that moral cognition, like much social cognition, is largely automatic, scaffolded by tacit affective processes and encapsulated from explicit conscious reasoning. Moral judgements are expressions of the outcomes of tacit emotional processes. Even protagonists in these debates who reject the sentimentalist interpretation agree that the cognitive processes involved in moral judgement remain largely tacit, and that explicit reasoning plays the role of *ex post facto* rationalization.

This consensus is generated by a theoretical framework with two features. Firstly moral cognition is tested by judgements about hypothetical scenarios. Secondly the judgements elicited are treated as the result of *either* explicit application of a moral rule or principle (maximize utility, treat people equally, don't use people as a means etc.) *or* a tacit cognitive or affective process. Experiments seem to show that in many cases tacit processes are essential to moral judgement while explicit reasoning is extraneous. Experimentalists then probe the nature of tacit cognitive processes involved using a combination of imaging and deficit methodologies standard in cognitive neuroscience. A particularly important set of results involves patients with damage to the ventromedial prefrontal cortex (more accurately the orbitofrontal cortex, directly behind the eyes) which affects tacit affective processes. These patients show atypical patterns of reinforcement learning, personal decision making, and moral judgement (Saver and Damasio 1991, Moll et al. 2002, Koenigs et al. 2007). The sentimentalist conclusion drawn from these experiments and others implicating affective processing in moral judgement is that moral judgement depends essentially on tacit affective processes. In the rest of this paper we refer to this version of sentimentalism as *neurosentimentalism* since it uses neuroscientific evidence to vindicate a sentimentalist position in metaethics.

We argue against neurosentimentalism on the basis that it divorces moral (and practical) judgement from moral (and practical) agency. This suggests neurosentimentalism is a far more radical doctrine than previously recognized. Neurosentimentalists took themselves to be vindicating the idea that moral judgement depends essentially on emotions, not the claim that moral judgement and moral agency are independent of each other. We take another approach to the evidence arguing that if moral judgement requires moral agency, then deficits in a capacity for moral judgement consequent on brain damage should be reconceptualized as impairments of agency. This does not require us to dismiss the evidence about the role of tacit emotional and cognitive processes in moral judgement but rather to re-evaluate their role in constituting the moral agent who makes judgements of rightness and wrongness.

The gap between judgement and agency created by neurosentimentalism emerges when we consider that the cognitive theories which provide the background to the neurosentimentalist interpretation of the performance of ventromedial prefrontal patients were developed in part to explain how people with amnesia could still make

preference-based judgements or perform categorization tasks using previously acquired information no longer explicitly available to them. The fact that these amnesics registered tacit emotional responses to preferred stimuli previously encountered is the basis of the theory that tacit affective processes can be the basis for decision-making.

Equally, amnesics who retained a capacity for short term explicit reasoning (at least moment to moment) could apply a rule to a case, even if they could not recall previous applications of the rule or store an occurrent rationale to use in future.

Thus the decision making options for patients with amnesia presented with a choice are either to reason explicitly about it using currently available information or to be prompted by tacit preferences. Their database of relevant previous experience has either been destroyed or is no longer retrievable.

Although ventromedial prefrontal patients are not amnesic, the neurosentimentalist interpretation of their decision making deficits reduces decision making to the exercise of the same restricted set of cognitive capacities available to an episodic amnesic. The ventromedial prefrontal cortex is standardly described as a mechanism for associating tacit affective responses with explicitly-represented information, thereby enabling the formation of preferences and decision making. When that associative process is inaccessible, due to ventromedial damage, the subject has to rely on explicit reasoning alone, producing abnormal judgements in many areas of personal decision making including moral deliberation. In standard experiments the relevant reasoning processes are taken to be the application of procedural rules of inference to semantic knowledge: drawing conclusions from propositions.

We shall call theories which take explicit procedural reasoning or tacit affective processing to exhaust the possible bases for judgement *synchronic* theories, since they do not require that the subject have explicit recall of previous events or be able to imagine the future consequences for herself of the judgements she makes. Instead she can either reason explicitly about the options using currently available information or act on a tacitly represented preference which surfaces in consciousness only as a gut feeling or inclination. Amnesics making judgements by either method are not moral agents, for familiar reasons, and cannot be held morally accountable for their judgements. Amnesics have no explicit access to a recalled or imaginatively constructed autobiographical history (we will discuss the evidence that amnesia impairs not only episodic memory but imagination). We shall

suggest that agency requires the capacity for episodic memory (of events in a subjective autobiography) and imaginative projection into the future in order for the subject to have the requisite intertemporal perspective on her actions.

Neurosentimentalism is a synchronic theory and our argument against it proceeds by showing that it cannot do justice to the connection between moral agency and moral judgement. A moral agent needs to be able to conceive of *herself* as a temporally extended entity as a necessary condition for moral reflection and decision-making. Yet the recent work in cognitive neuroscience, especially on patients with impaired ventromedial functioning, concentrates on synchronic judgements of rightness or wrongness of hypothetical actions, which do not require this type of intertemporal perspective on action. Those judgements can be made by applying a rule to a case which is why ventromedial patients do make impersonal judgements without experiencing the conflict produced in normal subjects by personal and emotional aspects of a situation.

We will argue that in fact there is good evidence that ventromedial damage impairs agency as well as judgement, and that the impairments in judgement may well result from impairments to agency.

Of course subjects with ventromedial prefrontal cortical damage do not have episodic amnesia, but we shall try and show that they do have impairments in the ability to construct a temporally extended autobiography, and that these impairments affect their capacities for agency in a similar way to amnesia. Thus, ultimately, our challenge to neurosentimentalism takes the form of a dilemma. Neurosentimentalism requires that *either* moral judgement does not require agency *or* amnesics and ventromedial patients can be moral agents. Neither alternative is attractive.

The rest of the paper proceeds as follows. In the first section we describe the dual process model of the mind which generates the contrast between tacit affective processes and explicit reasoning which shapes the synchronic theory. For dual process theorists moral judgement depends on distinct and dissociable systems. One cognitive system which recruits limbic and posterior neural structures performs 'social-emotional responses ... inherited from our primate ancestors ... shaped and refined by culture bound experience' (Greene et al. 2004, p. 398). As in all dual process theories these 'social-emotional' processes are tacit and automatic. Another cognitive system which recruits prefrontal cortical structures performs 'abstract thinking and high level cognitive control' (Greene et al. 2007, p. 398).

These explicit processes, as we shall call them, occur more slowly under the voluntary control of the agent.

In the second section we show how dual-process models support versions of neurosentimentalism and/or rationalist externalism (the view that moral judgements based on rational processes are not intrinsically motivating). The short argument is that tacit social-emotional processes are intrinsically connected to action while the explicit processes are motivationally inert. This view that reason is motivationally inert is an ancient philosophical one which now claims empirical support from cases of dissociation between motivation and explicit reasoning following damage to the ventromedial prefrontal cortex. We discuss the somatic marker hypothesis, which provides a conceptual framework for the application of dual process theories to moral decision-making and is now recruited by neurosentimentalists as empirical refutation of rationalist internalism.

In the third section we develop the agentic framework for interpreting the results which give rise to the somatic marker hypothesis and neurosentimentalism, arguing that the deficits in question impair agency not just judgement. Given that we are contesting the use of empirical evidence to support the synchronic interpretation we need to provide an empirically adequate diachronic account of agency. We suggest that crucial cognitive processes are episodic memory and imagination which combine under executive control to allow a person to conceive of herself as temporally extended (the capacity has been baptised mental time travel because it allows a self to escape the present moment by constructing and reconstructing alternative personal histories, linking episodes of memory and imagination). A striking feature of patients with ventromedial prefrontal damage is that they exhibit not only the unusual patterns of judgement emphasized by the synchronic conception, but impairments of agency and mental time travel. Our claim is not that these patients are amnesic (their damage is not to areas which cause episodic amnesia) but that their inability to use executive processes to assemble and inhabit the episodes necessary to support adequate deliberation impairs their capacity for agency and hence judgement.

Finally we consider the consequences of this wider interpretation for the sentimentalist/externalist picture that dominates the theoretical landscape. On our view synchronic versions of either theory are not plausible. Moral judgement cannot be reduced to automatic preferences evoked by the situation or, as rationalist externalists have

argued, the synchronic application of procedural rules. Our wider interpretation makes room for *diachronic* sentimentalists and rationalists who wish to incorporate a role for reflection and deliberation linked to a personal, intertemporal, perspective on action. The key is understanding the role played by the ventromedial prefrontal cortex in helping to create a moral agent constituted by the connection between deliberation, emotion, and diachronic selfhood, not just a decision calculus contingently connected to phylogenetically ancient motivational impulses.

## 1. Dual processing theories

Dual processing theories partition cognition according to the degree of conscious control a subject is able to exert over a process (Barbur et al. 1993, Weiskrantz et al. 1974). Conscious control is related to a number of other features, slow processing speed, voluntariness, effortfulness, and domain generality. These types of voluntary process are typically recruited in planning and deliberation and require the flexible control of attention and working and long term memory to retrieve, manipulate, compare, and evaluate representations of varying degrees of abstractness. We will call these processes *explicit* since the term connotes availability to higher, central, metacognitive processes required for voluntary control.

These explicit processes contrast with specialized cognitive processes which occur automatically beyond the intentional or conscious control of the agent. These automatic, involuntary, processes are typically rapid and inflexible. Importantly they include affective responses to stereotyped stimuli.

That affective processes can be tacit is shown in many neuropsychological conditions (e.g. Tranel and Damasio 1988a, b). Blindsight, amnesia, prosopagnosia, and hemi-neglect (see Faulkner and Foster 2002 for a review) provide cases in which tacit affective processing is preserved following impairment to explicit processing. Famously a densely amnesic patient acquired aversion to shaking hands with a person from whom she received a pinprick even though she could not remember the painful episode. A similar dissociation has been demonstrated for explicit and implicit *decision making*, leading to the somatic marker hypothesis popularised by Antonio Damasio and collaborators and applied to the case of moral cognition by sentimentalists such as Jonathan Haidt.

The somatic marker hypothesis is a theory of the role emotions play in decision making developed to explain decision-making deficits in brain-damaged patients with a lesion of the ventromedial prefrontal cortex (Damasio et al. 1991, Bechara et al. 1996). The Iowa Gambling Task was created to probe the nature of these deficits.

In the Iowa Gambling Task subjects, starting with a sum of money provided by the experimenter, draw 100 times from four decks of cards which prescribe a monetary outcome (e.g. receive \$100 from a bank provided by the experimenter or pay \$10 to the bank). Some contain cards which provide small gains and losses leading to a modest overall profit. Others contain cards which prescribe larger rewards but intermittent large losses leading to an overall loss. Typically neurotypical subjects start by choosing at random and then develop a preference for rewarding rather than punishing schedules. Importantly this pattern of advantageous selection develops before the subjects can explicitly articulate the punishment schedule as a rationale for their selections. Thus it appears that ability to learn and subsequently retrieve information about the value of the decks involves tacit processes.

Importantly the development of advantageous preferences is correlated with the presence of ‘anticipatory’ Skin Conductance Responses (SCRs) when considering decks associated with punishments. SCRs are considered to be indices of ‘somatic markers’: tacit affective responses to stimuli produced by the amygdala and limbic system. (The ‘somatic’ terminology reflects the James Lange theory of emotions which has it that emotions are ultimately bodily states.)

This finding is consistent with findings that subjects with impaired amygdala function show diminished SCR to emotionally-salient stimuli and fail to learn from experience, with disastrous consequences for decision making (Bechara et al. 1999). The same is true of patients with damage to the ventromedial prefrontal cortex, leading to the hypothesis that the ventromedial prefrontal cortex is a structure which allows subjects to access and use tacit affective memories in situations which require learning and decision making. This is reflected in their performance in the Iowa Gambling Task in which their choices do not come to reflect the punishment schedules even though they can learn them explicitly. Unlike neurotypicals, however, this explicit learning is not preceded by a period of implicit learning associated with SCRs or by the acquisition of a preference for rewarding decks.

Thus the patient with damage to the ventromedial prefrontal cortex is in the paradoxical situation of being able to understand

intellectually that some decks are disadvantageous but nonetheless choosing the disadvantageous decks.

Bechara explains their performance this way:

When [neurotypical] subjects decide to select from a specific deck the neural activity pertaining to this information is signalled to the ventromedial prefrontal cortex which in turn activates the amygdala (Damasio et al. 1991). This latter activity would reconstitute a state that integrates numerous conflicting instances of reward and punishment encountered with individual draws from that deck. If, in the end, negative somatic states outweigh the positive ones, an overall negative state is enacted and is indexed by the anticipatory SCRs we observed before the selection of cards from the disadvantageous decks. In turn, this influences the decision to avoid the deck under consideration. (Bechara et al. 1999, p. 5480)

To summarize: in the experiment neurotypical subjects are able to make choices which reflect previous patterns of reward and punishment *before* they can consciously describe those patterns or articulate a rationale for their choices. On a very popular dual processing interpretation of these results they can do so because the ventromedial prefrontal cortex retrieves tacitly represented information about the valence of options which then plays a role in regulating decisions (see Dunn et al. 2006 for a critical review). The crucial point is that this ‘trafficking’ role (Marc Hauser 2006) of the ventromedial prefrontal cortex is tacit and automatic. At the level of conscious processing the subject may only have a ‘gut feeling’ or be aware of a preference. This role for the ventromedial prefrontal cortex is shown by the fact that patients with ventromedial prefrontal cortex damage do not make advantageous decisions even though they can explicitly learn the punishment schedules.

These types of case are marshalled in support of two theses crucial to the current debate over the nature of moral cognition. The first is that explicit and tacit processes dissociate. The second is that tacit somatic markers which determine preferences need not be available to consciousness: they may be experienced as ‘gut feelings’ or inclinations. The subject cannot verbally explicate them or modify them using explicit reasoning processes. They are cognitively impenetrable as cognitive scientists say (Pylyshyn 1999).

## 2. Dual processing and metaethics

There are now many experiments and neuropsychological case studies demonstrating similar dissociations between explicit and implicit

cognition in moral judgement. Data collected by Jonathan Haidt suggests that the bulk of our ordinary moral judgements are the products of automatic unconscious processing. Haidt argues that moral judgements are ‘gut feelings or intuitions’ (Haidt 2002, p. 54). These ‘moral intuitions’ are a form of automatic process:

the sudden appearance in consciousness of a moral judgment, including an affective valence (good-bad, like-dislike), without any conscious awareness of having gone through the steps of searching, weighing evidence, or inferring a conclusion. (Haidt 2001, p. 818)

According to Haidt moral intuitions normally lead directly to moral judgements, explicit moral reasoning is rare, and if moral reasoning *is* engaged it is ‘not left free to search for the truth but is ... employed only to seek confirmation of preordained conclusions’ (Haidt 2001, p. 822). Haidt claims on the basis of a series of ‘moral dumbfounding’ experiments in which subjects were left unable to justify their moral responses to, for example, sibling incest, that moral reasoning is merely post hoc rationalization of emotionally driven responses. In these experiments subjects were provided with an incest scenario from which the common justifications for condemning incest were removed—there was no possibility of pregnancy, the siblings’ relationship was not damaged, etc.—yet they showed no disposition to revise their negative judgements when this was pointed out to them.<sup>1</sup> Haidt concludes that the roots of morality are to be found not ‘in our ability to search and evaluate evidence in an open and unbiased way’ using explicit reasoning but rather ‘in what the mind [tacitly] does best: perception, intuition, and other mental operations that are quick, effortless, and generally quite accurate’ (Haidt 2001, pp. 821–2).

The meta-ethical positions which are most clearly accommodated by a dual processing model are versions of sentimentalism and externalism. Indeed Haidt explicitly takes his work to have demonstrated Hume’s sentimentalist claim that reason is always and only a slave to the passions. Sentimentalists think that moral judgements are essentially expressions of emotional responses to situations, acts, or agents driven from below. ‘When we utter “ought”’, says Jesse Prinz, ‘we express our own sentiments and factual knowledge is not sufficient for having sentiments’ (2006, p. 38). One could not, he argues, ‘sincerely attest that killing is wrong without being disposed to have negative emotions towards killing’ (2006, p. 32). Neurosentimentalism

<sup>1</sup> Haidt’s results have been widely discussed and his interpretations of them are not uncontroversial. (See e.g. Saltzstein, D. Herbert, and Tziporah Kasachkoff 2004.)

would be largely vindicated if moral judgement depended on cognitively impenetrable emotional processes, and in fact Prinz and others explicitly argue from evidence that moral judgement depends on processes, typically involving the affective systems, which are not sensitive to beliefs, expectations, and goals, to sentimentalism. Prinz claims that emotions are both *necessary* and *sufficient* for moral judgement: he points to work by Wheatley and Haidt suggesting that ‘a negative feeling can give rise to a negative moral appraisal without any specific belief about some property in virtue of which something is wrong’.<sup>2</sup> If gut feelings are both necessary and sufficient for moral judgement then it seems clear that the neurosentimentalist account of moral judgement does not require a moral agent.

The externalist account of moral judgement on the other hand locates it entirely in controlled, reflective processes. Claims about what is morally required are to be thought of as judgements about the correct application of a set of moral standards to some situation, arrived at through explicit procedural reasoning. Moral judgements are argued to be thus analogous to judgements about the requirements of etiquette (e.g. Foot 1972) or of the law (Roskies 2003 and 2008). Our beliefs in these domains have no necessary connection to motivation. They can only motivate when paired with our contingent affective responses via the kinds of mechanisms described in the previous section. And it is conceivable on this account that agents with full moral knowledge (semantic knowledge of the standards and the conditions under which they apply) might be impaired either in the trafficking mechanism or in their affective responses and so fail to be motivated in accordance with their moral judgements. While some externalists might be willing to concede that automatic affective processes are necessary for the *acquisition* of moral concepts (perhaps in something like the way suggested by Shaun Nichols)<sup>3</sup> they maintain they are not essential to the *application* and manipulation of those concepts in

<sup>2</sup> Wheatley and Haidt 2005. In this experiment subjects were hypnotized to feel disgust when they heard morally neutral terms such as ‘often’. Subjects were more inclined to judge morally admirable characters negatively in vignettes containing the target word. However, as Jill Craigie has pointed out to us (personal communication), an examination of the data shows the effects were relatively slight.

<sup>3</sup> Shaun Nichols argues that the normal affective response to harm infuses core moral rules (those which specify the harm based violations) with a special status (Nichols 2002 and 2004). With regard to the psychopath Nichols argues that ‘since psychopaths have a deficiency in their affective response to harm in others, this plausibly explains why they fail to treat harm norms seriously’ (Nichols 2004, p. 300). But it is not clear that this generalizes to a broader set of moral rules.

moral reasoning and judgement. They thus interpret cases of ‘acquired sociopathy’ following damage to the ventromedial prefrontal cortex as demonstrating a dissociation between moral judgement and moral motivation.<sup>4</sup> We suggest that they are likewise committed to regarding amnesics, who retain semantic knowledge of moral rules and the capacity to apply them using currently available information, as full moral agents.

Insofar as the term ‘moral judgement’ is used to refer to an agent’s beliefs about the content of morality, conventionally understood, sentimentalists will agree that such judgements have no essential connection to motivation. And they will also agree that because emotional responses are independent of procedural reasoning the *rational* aspect of moral judgement is motivationally inert.

Sentimentalists and rationalist *externalists* thus disagree on where in cognition moral judgement occurs but they agree on the cognitive impenetrability of the automatic processes which provide motivation. The dissociation between intact reasoning and impaired affective processing evidenced by ventromedial patients can be recruited by sentimentalists or externalists because in effect they embrace different horns of a dilemma about the basis of moral judgement: procedural rationality or affective processing.

Rationalist *internalists* hold that the link between moral judgement and motivation is necessary. They therefore agree with sentimentalists that externalism does not provide a sufficient account of moral motivation. However, because rationalists think that genuine moral judgements are not simply emotional responses but must be the product of, or shaped by, reason, they require a different account of the connection between moral judgement and motivation than that offered by sentimentalists. The account of moral reasoning given by

<sup>4</sup> Adina Roskies (2003) relies upon the case of EVR to argue that patients with ventromedial frontal lobe damage possess unimpaired moral reasoning abilities, yet are not motivated by the moral judgements they make. EVR had resections of his ventromedial prefrontal cortex in order to remove a brain tumour. He subsequently attracted a diagnosis of ‘acquired sociopathy’ on the basis of features such as an ‘inability to sustain consistent work behavior’, ‘lack of ability to function as a responsible parent’, and ‘defective planning’. According to Roskies this, plus his failure to produce anticipatory skin conductance responses on the Iowa Gambling Task and another task, is evidence of a failure of moral motivation. However EVR performed at an advanced level on a Kohlbergian moral reasoning test (Saver and Damasi 1991) and gave principled responses to hypothetical ethical dilemmas (Eslinger and Damasio 1985) and Roskies claims this as evidence that his capacity for moral judgement is unimpaired. For discussion of these claims see Kennet and Fine 2008a, 2008b, and Roskies 2008.

externalists which relies on narrowly rational processes of inference from a set of rules or principles — processes which may be intact in patients with amnesia and ventromedial damage — is clearly insufficient to deliver this link.

The insufficiency of affective processing and procedural reasoning (even when they are both intact as in amnesia) results from the role played by these processes in the life of the subject. A person, such as an amnesic, who cannot incorporate procedural judgements or affective preferences *as her own* is not an agent. The reason is that she cannot represent the results of these processes to herself, because she is not a diachronic self but a bundle of habits linked to a synchronic reasoning system.

There are a number of rationalist stories about the deliberative processes that underpin moral judgement and the norms which govern them. In all these accounts moral deliberation and judgement cannot be separated from moral agency.

So for example, Michael Smith argues that ‘it is a conceptual truth that claims about what we are morally required to do are claims about our reasons’ (Smith 1994, p. 84). Smith here is referring to *normative* reasons; considerations which individuals take to guide practical reflection and to justify, rather than merely passively explain, their decisions. On this account for an agent sincerely to attest that killing is wrong is for that agent to accept that she *herself* has a (decisive) reason not to kill and this has direct practical implications for her choices and actions. For rationalists therefore, moral judgements are exercises of agency. Indeed Christine Korsgaard claims that it is ‘from the standpoint of practical reason that moral thought and moral concepts ... are generated’ (Korsgaard 1988, p. 132).

On the rationalist internalist interpretation then, patients with ventromedial damage and amnesia either fail to make such judgements, or are impaired *as agents* in their inability to choose and act in accordance with reasons they accept.

But this rationalist position cannot be mapped onto the dual process model which sees action as either driven from below by automatic affective preferences or from above by procedural reasoning detached from affect and personal motivation. And more sophisticated forms of sentimentalism which require reflective endorsement of the deliverances of automatic affective processes are likewise not fully represented by a dual process model since these more sophisticated diachronic sentimental accounts also require that our tacit emotion based judgements be cognitively penetrable — moderated by

deliberation and reflection and sensitive to *our* beliefs, plans, and expectations. Does the dual process model show that these positions are empirically unrealizable? Or should we rather suspect that the model provides an incomplete picture of human moral cognition by failing to adequately describe the processes of practical reflection which are central to these views? In what follows we explore a process which we argue is critical to practical reflection, to self-regulation, and so to the construction of moral agency itself.

### 3. Mental time travel and myopia for the future

Antoine Bechara coined the term ‘myopia for the future’ to describe the impairment introduced by damage to the ventromedial prefrontal cortex. Although these patients know that some decks in the Iowa Gambling Task are disadvantageous they still choose from them. This is not because they are indifferent to punishment. Bad decks produce aversive SCRs at the time of selection. Patients with damage to the ventromedial prefrontal cortex however do not produce ‘anticipatory’ SCRs when contemplating disadvantageous choices. Hence the idea that damage to the ventromedial prefrontal cortex introduces an inability to *foresee the consequences* of adverse decisions.

Most accounts of the role of the ventromedial prefrontal cortex in decision-making do not provide an account of foresight per se. Rather they suggest that foresight reduces to *either* explicit reasoning using semantic knowledge, such as expected utility calculation, *or*, in a well-calibrated mind, explicit reasoning reinforced by tacit emotional processes.

However there is another way to understand foresight, concentrating on the intimate relationship between memory and imagination in planning. Recent work on memory has concentrated on the role in decision-making of episodic memory. Episodic memories are those in which a subject recreates the experience of a previous episode in her life. There seems to be a subtle difference between ‘pure’ episodic memory, which recreates the perceptual or sensory content of experience, and autobiographical memory. In the latter episodic memories combine with a sense of self, of being personally present in the episode (known as autoneoasis) (Klein, Loftus, and Kihlstrom 2002, Levine et al. 1999, Levine et al. 2004, Tulving 2002, and Wheeler et al. 1997). Autoneoasis is a term of art intended to capture that aspect of self consciousness which annexes experience to the self not just at a

time, but over time. We might describe it as awareness of diachronic selfhood.

Of course in the normal case of episodic memory recruited for planning the ‘mine-ness’ of an episode is experienced together with its content. My episodic memory of a sunny day at the beach is experienced as mine (Steinvorth et al. 2006). However it seems that ‘mine-ness’ is a cognitive achievement mediated by the ventromedial prefrontal cortex. While ‘pure’ episodic memory studies (such as recall of visual scenes) do not activate the ventromedial prefrontal cortex, ‘activations of the ventromedial prefrontal cortex are almost invariably found in autobiographical memory studies’ (Gilboa et al. 2004, p. 1336). Gilboa suggests that this is because ‘autobiographical memory relies on a quick intuitive “feeling of rightness” to monitor the veracity and cohesiveness of retrieved memories in relation to an activated self-schema’ (Gilboa et al. 2004, p. 1336). Our interpretation of these results is that episodic memory is normally recruited for planning, which is coordinated by the ventromedial prefrontal cortex, and that planning is the activity of a diachronic self. Thus autobiographical memory is episodic memory annexed to a self in the service of executive activity.

Recent accounts of episodic memory have shifted in emphasis from distinguishing it from semantic memory (using paradigms such as the ‘know/remember’ test in which memory of either type is detached from personal goals—the targets are typically word lists or pictures) to understanding the contribution of autobiographical memory to global executive functions such as planning. However once we think of episodic memory this way, as providing autobiographical episodes for planning purposes, it seems that the encoding and retrieval of *previous* experience is only part of the story. Planning and decision-making involve information about the *future* as well as the past. And, as with memory, this information can be semantic or episodic. As well as representing factual knowledge about the future we can imagine ourselves living out future scenarios, rehearsing different possibilities. This form of imaginative rehearsal is known as *prospection*: the future-directed analogue of episodic memory (Suddendorf and Corballis 1997 and 2008).

And indeed it seems that planning involves the seamless integration of both forms of projection of oneself in time, episodic memory and *prospection*. For this reason the ability to retrieve past episodes and imagine future ones and to integrate the results with other forms of knowledge as part of planning have been baptized *mental time travel*

(Bayley et al. 2003, Suddendorf and Corballis 2008, Suddendorf and Busby 2005, D'Argembeau and van der Linden 2006, Buckner and Carroll 2007, and Buckner et al. 2008).

There is an emerging body of evidence that mental time travel does not exploit different systems for memory and imagination. Rather both memory and imagination involve the activation of relevant perceptual, sensory, and emotional systems in the absence of an environmental stimulus (Hassabis et al. 2007, Klein et al. 2002, and Miller and Cohen 2001). Thus the essential feature of mental time travel is the ability to create and recreate these experiences under voluntary control rather than via the presentation of an eliciting situation or object (Schacter, Addis et al. 2007 and 2008). The consequences for planning are enormous. Mental time travel gives humans a database of situations and responses to them which can be safely rehearsed offline. It is not a coincidence that amnesic patients such as Alzheimers cases are deficient in planning and executive function. Their problem is not just loss of memory per se but the consequences for executive action (Buckner et al. 2008).

It is this voluntary, executive, aspect of mental time travel which is so important for the capacity for planning, and which makes executive action dependent on maturation of the frontal systems (Frith 1996, Gerrans 2007, S. Kapur et al. 1995, Knight 1999, Miller and Cohen 2001, Waltz et al. 1999, Wood and Grafman 2003). Episodic information is encoded in a distributed system. Prefrontal systems coordinate its retrieval and manipulation. We might say that there is a distributed episodic database encoded by specialised systems<sup>5</sup> which can be accessed and manipulated by frontal systems involved in mental time travel to create actual and possible autobiographies. This is why:

[p]atients with damage restricted to the frontal cortex are impaired when mentally required to re-experience a study episode in sufficient detail to recollect contextual information about that episode even though they can often report about the factual contents of the same episode. (Wheeler et al. 1997, p. 342)

If this is the case then mental time travel could be compromised at different levels, depending on whether damage was to the episodic database *or* to the frontal systems which access and manipulate the

<sup>5</sup> Of course we do not think that this database is localized. The hippocampus and medial temporal lobes, though a memory system, do not directly encode percepts. Rather they encode traces which allow the reconstruction of the relevant representations as required by current cognitive contexts. 'Episodic database' is shorthand for this conception of memory.

data in executive processes. Classic cases of amnesia correspond to the former and frontal damage which leads to failures of executive retrieval and manipulation of autobiographical information to the latter.

It is for this reason that recent work on the role of frontal systems in mental time travel is suggestive. It seems likely that the normal case of decision-making involves mental time travel. That is to say the subject remembers what happened last time and uses that information to create and inhabit a future scenario. In contrast a patient with ventromedial damage cannot perform mental time travel. Not because she lacks an episodic database but because she cannot make use of it.

This hypothesis predicts that patients with ventromedial damage would perform poorly on a mental time travel task since such patients would have an intact experiential database but impaired ability to retrieve or manipulate it. Unfortunately the performance of patients with ventromedial damage in mental time travel tasks has not been directly tested, although to the extent that decision-making tasks depend on mental time travel the results are consistent with the prediction. Nonetheless Levine et al. (1999) report a case of a patient, M.L., with amnesia and ventromedial damage with deficient mental time travel, as one would predict, *and* decision making deficits characteristic of Self Regulation Disorder which indicates a loss of agency.

A second prediction is that patients with a deficit in mental time travel would perform poorly on the Iowa Gambling Task irrespective of their ability to generate SCRs to disadvantageous decks.

Gutbrod et al. (2006) tested the performance of amnesic patients on the Iowa Gambling Task. Clearly such patients are disabled with respect to mental time travel since they lack the database. Interestingly, 9 of 11 patients performed at chance and did not show differential anticipatory SCRs to advantageous and disadvantageous decks. Furthermore the magnitude of anticipatory SCRs did not correlate with behavioural performance, leading to the conclusion that ‘acquisition of a behavioural preference—be it for advantageous or disadvantageous choices—depends on the memory of previous reinforcements encountered in the task, *a capacity requiring explicit memory*’ (Gutbrod et al. 2006, p. 1315. Emphasis added). Of course the explicit memory involved may be semantic rather than episodic, but that explicit memory proves necessary for learning is as troubling for the somatic marker hypothesis as evidence that somatic marking is not always necessary (Heims et al. 2004).

The fact that amnesics with an intact capacity for SCR perform poorly in the Iowa gambling task is a contribution to the debate

over whether or not explicit or implicit learning is involved in Iowa Gambling Test performance (Maia et al. 2004 and 2005). It is also evidence in favour of a richer interpretation of the nature of the deficits in ventromedial cases because it suggests that a deficit in mental time travel can lead to a deficit in decision-making.

The nature of that deficit is important. It is not just loss of information about future consequences but of an essentially indexical way of representing that information which impairs deliberation.

#### 4. Mental time travel, agency, and morality

It is uncontroversial that only individuals who meet certain threshold conditions for agency can be morally accountable for the judgements and decisions that they make, and it is surely the analysis of the moral judgements of *those* individuals that is the primary task of meta-ethical accounts of moral judgement. We suggest here that the capacity for mental time travel is critical in constituting the person as an agent capable of responding to moral demands. Michael Bratman (2000) argues that our reflectiveness, our planfulness, and our conception of our agency as temporally extended are core interrelated features of human agency. We do not simply act from moment to moment. We are not trapped in an eternal present. Rather, Bratman claims, we conceive of ourselves as agents who persist over time and so we construct and commit ourselves to future directed plans, which we intend should structure and coordinate our more particular decisions and activities. Such acts of planning and commitment clearly require cognitive resources that go beyond the pairing of tacit responses with explicit understanding of rules or outcomes.

Planning requires a capacity to imaginatively project oneself into the future; this in turn requires both a sense of oneself as the very same individual who will inhabit that future (autonoetic awareness), and also the kind of detailed self-knowledge that is supported by autobiographical memory. We exercise the capacity for mental time travel whenever we revise for this year a class we gave last year — remembering what worked and what didn't — whenever we reflect on what kind of career or job would best suit us, whenever we plan a holiday or a shopping trip, arrange a meeting, organize a party, or commit ourselves to a course of study, an exercise program, or a marriage. In so committing ourselves, as Velleman (1991 and 1997) points out, we provide reasons for ourselves in the future, reasons

which will be *ours*, but which we would otherwise not have had. In this way we construct ourselves as particular, temporally extended, *agents*. Our diachronic reasons, made salient to us via our capacity for mental time travel, are thus in a position to compete with synchronically occurring wants. In effect they become normative for us.

The capacity for mental time travel is arguably required in order for *any* demand to be normative for a person — for them to have reasons whose force is independent of immediate stimulus bound responses. Morality, as Jay Wallace (1999) argues in his defence of internalism, is very widely taken to be a normative domain. We suggest that the process of becoming an agent capable of engaging in this normative domain just is the process of learning to transcend the present moment, both cognitively and behaviourally. This necessarily involves the meta-cognitive capacity to conceive of and adopt reasons which extend over time. In this respect we are in sympathy with Christine Korsgaard (2002) when she says that reasons must be defined in terms of our rational nature, rather than the other way round. She asks what is it about us that makes reasons exist for us, and answers that it is our self-consciousness. Perhaps surprisingly then, this version of rationalism finds empirical support in some of the central cases neurosentimentalists and externalist draw upon, and we think better explains our ordinary intuitions about the absence or impairment of agency across a range of cases.

Insofar as genuine moral judgements must be made by those who meet a threshold of moral agency, and such agency depends in part upon the capacity for mental time travel, this capacity is necessary for moral judgement. The recognition of the role played by mental time travel in practical reasoning and agency helps to explain why animals and small children, who may display the tacit affectively based responses (and in the case of small children some explicit grasp of moral rules) highlighted in neurosentimentalist accounts of moral judgement, are not counted as full moral agents. It also explains why they are not held morally responsible for their actions, as well as having significant implications for the interpretation of moral judgement in cases of amnesia and ventromedial damage. Moral agents must be able to take some considerations as *normative* and this we suggest is what those who are impaired in the capacity for mental time travel or in whom the capacity is not developed, cannot do.<sup>6</sup>

<sup>6</sup> Our view finds further support in studies of the self-regulatory capacities of children which have found links between levels of inhibitory control (which includes the rules of delay)

Of course there are a number of different ways in which an adult person's agency may be impaired, not all of which turn upon failures of mental time travel. An individual is a grossly impaired agent if the plans and decisions she arrives at in deliberation are constantly overwhelmed by compulsive desires or if she is so depressed that she cannot bring herself to act at all. In these kinds of cases there may be no failure of the capacity for mental time travel. But other cases of impaired agency do turn upon failures to conceive of or feel appropriately connected to one's future self. An individual will not meet the threshold for moral agency if she cannot make choices and decisions which commit her future self to certain courses of action, either because she is incapable of delving into her personal future and evaluating future possibilities or because she will not, when the time for action arrives, be able to remember or feel normatively bound by the plans she makes now. Where the capacity for mental time travel is seriously damaged there is no sense in which the individual might be said to be shaping her own life or acting on the basis of reasons over enough time so as to satisfy the threshold for moral agency.

Patient M. L. reported by Levine et al. appears to be such a case. M. L. was unable to episodically re-experience post-injury events to the same extent as control subjects, although he could use familiarity or other non-episodic processes to distinguish events he had experienced from those he had not experienced. He continued to report a feeling of *subjective distance* from recall of events occurring *after* his recovery. His errors of judgement and failures to understand his responsibilities as a parent required supervision of his behaviour and structured routines, suggesting that merely semantic access to past and future events is not enough to support agency. A more profound case of amnesia is reported by Tulving (1985). Patient N. N. could not recall a single instance from his past nor imagine the future. When asked what he would be doing tomorrow he would reply that he did not know. When asked to characterize what it is like for him when he tries to think about the past or future he described it as 'like being asleep' or 'it's a big blankness', 'It's like being in a room with nothing there and having a guy tell you to go find a chair and there's nothing there' (Tulving 1985, p. 4). That patients whose agency is so grossly

---

and moral development (Kochanska et al. 1996 and Kochanska et al. 1997). Children with high levels of control—that is, those who are successful in extending their agency across time—show higher levels of internalization of normative rules of conduct. Moreover, different levels of inhibitory control predict subsequent moral development.

impaired might nevertheless be able to satisfy neurosentimentalist or externalist accounts of what it takes to make moral judgements, as we have suggested they might, calls for a response from proponents of those accounts.

Ventromedial patients too are notably impaired in their capacity for planning and here too we think this affects their capacity to take moral considerations as genuinely normative. The typical ventromedial patient is described as impulsive and unable to plan, and incapable, because unmotivated, of acting on the basis of conclusions she reaches as a result of rational deliberation. They appear uninterested in their own shortcomings. Such a person might acknowledge for example that money is required for a spouse's operation but spend the money on an impulse. She might then agree that the money should have been spent on a spouse's operation, but show no remorse or real interest in the matter.

Antonio Damasio has described such patients as having intact rational capacities but lacking the ability to associate those rational processes with emotionally significant information. For such a person explicit reasoning about even personally significant events becomes motivationally inert. He and others sometimes speak as if what is lost by ventromedial patients is the ability to associate explicit semantic representations such as linguistic thoughts, or the reasoning processes which manipulate them, with emotional valence, and hence to act appropriately on the results of rational deliberation.

What is distinctive about this story is the poverty of its characterization of the process of rational deliberation and indeed of planning. It leaves out the fact that choices are made and upheld in an autobiographical context generated by mental time travel, not just via the association of hedonic states with options to generate a preference ranking. The framework of autobiographical memory gives sense and meaning to the activities of deliberation and choice, which may be considerably more nuanced, sophisticated, and meaningful than the simple generation of future oriented gut reactions.

An alternative interpretation of frontal lobe patients which focuses on the capacity for mental time travel is provided in a review article by Wheeler, Stuss, and Tulving (1997). They associate frontal pathology and the decision making impairments observed in such patients with deficits in recollection of the past, introspection of the present, and foresight, which they interpret as disruptions in auto-noetic consciousness. The problem as they see it is that affected patients have difficulty in relating available personal information to themselves. While such

patients know about things that have happened to them they have lost the sense of personal connectedness to those events and so cannot make appropriate use of their knowledge in deliberation.

The apparent dissociation in patient EVR, an exemplary ventromedial patient, between superior third-personal hypothetical reasoning and judgement and grossly impaired capacity for first-personal planning and decision making, is particularly suggestive of an inability to inhabit scenarios in his personal future.

Deciding where to dine might take hours, as he discussed each restaurant's seating plan, particulars of menu, atmosphere, and management. He would drive to each restaurant to see how busy it was, but even then he could not finally decide which to choose (Eslinger and Damasio 1985, p. 1732).

And in referring to tests of social cognition during which he had to come up with solutions to social problems, Saver and Damasio report that 'EVR noted with his customary insight that he came up with many options but "...I still would not know what to do"' (Saver and Damasio 1991, p. 1246).<sup>7</sup> As in the case of amnesics his personal future is something of a mystery to him. It is not something over which he has agentive authority.

Mental time travel is the mechanism by which we acquire the phenomenology of a temporally extended self with an integrated past, present, and future. What ventromedial patients lose is not just the ability to plan but the prerequisite sense of diachronic selfhood intimately connected with an autobiography with a distinctive emotional character. This provides an explanation of the observed incapacity of ventromedial patients to translate their impersonal social judgements—the products of explicit reasoning—into personal practical judgements. They have lost the sense of the 'I' to whom the judgement is to be indexed so they literally don't know what to do.

## 5. Conclusion

Clearly claims of the form, *x is wrong/right*, can be produced by either tacit processes alone, explicit processes alone, or by iterative

<sup>7</sup> A patient reported by Stuss (1991) appears to show a similar dissociation between third-personal knowledge and first-personal judgements. When asked to assess someone's (her own) significant problems at work from a third-personal perspective (that of a supervisor), she was able to analyse the situation competently and make appropriate recommendations. Yet asked for similar judgements from a first-personal perspective, she refused to agree with her previous recommendations.

interaction between tacit and explicit processes. The question for meta-ethics is whether the kinds of judgements focused on by externalists and neurosentimentalists and produced by one or both of these cognitive processes deserve, or best deserve, the title of moral judgement.

We take ourselves to have shown here that it is a problem for sentimentalism and externalism that they make no essential connection between moral judgements and moral agents. For all they say the question of whether some individual has made a moral judgement is entirely disconnected from the question of whether they are a moral agent or an agent at all. But if we accept that moral judgements are essentially normative this cannot be the case, since such judgements can be normative only for moral agents. And this we have claimed, at least in human agents, requires in addition to tacit and explicit processing, the capacity for mental time travel.

In our view then, neurosentimentalism and externalism fail on empirical and conceptual grounds to offer an adequate account of moral judgement, because they leave the moral agent out of the picture. There is more to moral agency than the making of judgements using either tacit affectively scaffolded practices or the explicit application of rules. Our claim is that the dual process framework relied on explicitly by neurosentimentalists and implicitly by externalists, does not have the resources to distinguish between those who meet the threshold for moral agency and those who do not. Recourse to this framework in meta-ethics therefore results in a partitioning of moral judgement from the rest of the moral and practical domain or commits those who use it to the proposition that even densely amnesic patients can be moral agents.

If, as we contend, moral judgements require moral agents and agency is constituted in part through mental time travel this leaves unscathed those versions of rationalism and sentimentalism which allow a central place in moral deliberation and judgement to diachronic selfhood.

Our project has been to show that mental time travel is necessary to moral agency and so is a prerequisite for moral judgement, not that it is sufficient. As such it leaves the question of what else might be required for adequate moral judgement to further meta-ethical debate and empirical investigation. Sentimentalists will focus on empathic transference of concern from self to other. And it is likely that the processes used in mental time travel are also recruited when we imagine others' situations. Rationalists require that we be able to align

our personal planning and decision making with the results of explicit impartial forms of reasoning to override, where necessary, the influence of automatic tacit processing.<sup>8</sup> But for rationalists who take a constructivist view of moral agency, such as Korsgaard, there is no fundamental tension between moral and practical judgement. She says:

The territory of practical reasons is not split into two domains—self-interested rationality concerned with the occupant of this particular body on the one hand, and reasons of impartial morality on the other. Instead the personal concern which begins with one's life in a particular body finds its place in ever widening spheres of agency and enterprise, developing finally into a personal concern for the impersonal. (Korsgaard 1988, p. 127)

Finally we note that mental time travel does also play a significant and *direct* role in moral judgement. There are many cases of moral judgement which are not well captured by the synchronic focus of most experimental work on moral cognition, where automated decision making or rule application does not capture the phenomenology of moral decision-making and will not be adequate to the situation faced by the agent. Moral reflection, planning, and choice is often undertaken in an autobiographical context. It engages our sense of self and our capacity to see ourselves, and others, and the world in which we find ourselves, diachronically. We thus need to distinguish between the largely automated synchronic application of clear cut ingrained moral rules, such as those against killing and assault, and more difficult, nuanced, and complex decisions, such as whether to care for one's demented father at home or place him in care, whether or not to end a marriage, or to leave one's family behind in a refugee camp and take the only available sponsorship to emigrate to a new country. It is cases like these that press cognitive theories of morality to attend to the processes which create and sustain agency.

## References

Barbur, J. L., J. D. G. Watson, R. S. J. Frakowiak, and S. Zeki 1993: 'Conscious Visual Perception Without V1'. *Brain*, 116, pp. 1293–302.

<sup>8</sup> See Fine (2006) and Kennett and Fine (2009) for discussion of evidence that explicit reasoning may override automated responses and that reasoned responses may become automated over time.

- Bayley, P. J and L. R. Squire 2003: 'Successful Recollection of Remote Autobiographical Memories by Amnesic Patients with Medial Temporal Lobe Lesions'. *Neuron*, 38, pp. 135–44.
- Bechara, A., D. Tranel, H. Damasio, and A. R. Damasio 1996: 'Failure to Respond Autonomically to Anticipated Future Outcomes Following Damage to Prefrontal Cortex'. *Cerebral Cortex*, 6(2), pp. 215–25.
- Bechara, A., H. Damasio, and A. R. Damasio 1999: 'Different Contributions of the Human Amygdala and Ventromedial Prefrontal Cortex to Decision-making'. *Journal of Neuroscience*, 19, pp. 5473–81.
- Blair, R. J. R. 2004: 'The Roles of Orbital Frontal Cortex in the Modulation of Antisocial Behavior'. *Brain and Cognition*, 55, pp. 198–208.
- 2005: 'Applying a Cognitive Neuroscience Perspective to the Disorder of Psychopathy'. *Developmental Psychopathology*, 17, pp. 865–91.
- Bowman, C. H., C. E. Evans, and O. H. Turnbull 2005: 'Artificial Time Constraints on the Iowa Gambling Task: The Effects on Behavioural Performance and Subjective Experience'. *Brain and Cognition*, 57, pp. 21–5.
- Bratman, M. 2000: 'Reflection, Planning, and Temporally Extended Agency'. *Philosophical Review*, 109, pp. 35–61.
- Bruyer, R. 1991: 'Covert Face Recognition in Prosopagnosia: A Review'. *Brain and Cognition*, 15, pp. 223–35.
- Buckner, R. and D. Carroll 2007: 'Self Projection and the Brain'. *Trends in Cognitive Sciences*, 11, pp. 49–57.
- Buckner, R., J. R. Andrews-Hanna, and D. Schacter 2008: 'The Brain's Default Network. Structure Anatomy and Relevance to Disease'. *New York Academy of Science*, 1124, pp. 1–38.
- Busby, J. S. and T. Suddendorf 2005: 'Recalling Yesterday and Predicting Tomorrow'. *Cognitive Development*, 20, pp. 362–72.
- Christianson, S., A. Forth, L. Lidberg, R. Hare, and Lars-Hakan Thorell 1996: 'Remembering Details of Emotional Events: A Comparison Between Psychopathic and Nonpsychopathic Offenders'. *Personality & Individual Differences*, 20(4), pp. 437–43.
- Clayton, N. S., T. J. Bussey, and A. Dickinson 2003: 'Can Animals Recall the Past and Plan for the Future?' *Nature Reviews Neuroscience*, 4, pp. 685–91.
- Cullity, G. and B. Gaut (eds) 1997: *Ethics and Practical Reason*. Oxford: Clarendon Press.

- D'Argembeau, A. and M. Van der Linden 2006: 'Individual Differences in the Phenomenology of Mental Time Travel: The Effect of Vivid Mental Imagery and Emotion Regulation Strategies'. *Consciousness and Cognition*, 15, pp. 342–50.
- Damasio, A. R., D. Tranel, and H. Damasio 1991: 'Somatic Markers and the Guidance of Behaviour: Theory and Preliminary Testing'. In Levin, Eisenberg, and Benton 1991, pp. 217–29.
- Davies, M. 1987: 'Tacit Knowledge and Semantic Theory'. *Mind*, 96, pp. 441–62.
- Delazer, M. and T. Benke 1999: 'Arithmetic Reasoning and Implicit Memory'. *Brain and Cognition*, 40, pp. 94–7.
- Dunn, B. D., T. Dalgleish, and A. D. Lawrence 2006: 'The Somatic Marker Hypothesis: A Critical Evaluation'. *Neuroscience and Biobehavioural Reviews*, 30, pp. 239–71.
- Eslinger, P. J. and A. R. Damasio 1985: 'Severe Disturbance of Higher Cognition After Bilateral Frontal Lobe Ablation: Patient EVR'. *Neurology*, 35(12), pp. 1731–41.
- Faulkner, D. and J. K. Foster 2002: 'The Decoupling of "Explicit" and "Implicit" Processing in Neuropsychological Disorders: Insights into the Neural Basis of Consciousness?' *Psyche*, 8.
- Fine, C. 2006: 'Is the Emotional Dog Wagging Its Rational Tail, or Chasing It?' *Philosophical Explorations*, 9, pp. 83–98.
- Foot, P. 1972: 'Morality as a System of Hypothetical Imperatives', in her 1987, pp. 157–73. Originally published in *Philosophical Review*, 81, pp. 305–15.
- 1987: *Virtues and Vices and Other Essays in Moral Philosophy*. Oxford: Basil Blackwell.
- Frith, C. D. 1996: 'The Role of the Prefrontal Cortex in Self-Consciousness: the Case of Auditory Hallucinations'. *Proceedings of the Royal Society of London*, 351(B), pp. 1505–12.
- Gerrans, P. 2007: 'Mechanisms of Madness. Evolutionary Psychiatry Without Evolutionary Psychology'. *Biology and Philosophy*, 22, pp. 35–56.
- Gilboa, A., G. Winocur, C. L. Brady, and S. J. Hevenor 2004: 'Remembering our Past: Functional Neuroanatomy of Recollection of Recent and Very Remote Personal Events'. *Cerebral Cortex*, 14(11), pp. 1214–25.
- Greene, J., L. Nystrom, A. Engell, J. Darley, and J. Cohen 2004: 'The Neural Bases of Cognitive Conflict and Control in Moral Judgment'. *Neuron*, 44, pp. 389–400.

- Greene, J. and J. Haidt 2002: 'How (and where) Does Moral Judgment Work?' *Trends in Cognitive Science*, 6, pp. 517–23.
- Gutbrod, C., C. Kroužela, H. Hofera, R. Mürria, W. Perrig, and R. Ptak 2006: 'Decision-making in Amnesia: Do Advantageous Decisions Require Conscious Knowledge of Previous Behavioural Choices'. *Neuropsychologia*, 44, pp. 1315–24.
- Haidt, J. 2001: 'The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment'. *Psychological Review*, 108(4), pp. 814–34.
- 2002: "'Dialogue Between My Head and My Heart": Affective Influences on Moral Judgment'. *Psychological Inquiry*, 13, pp. 54–6.
- Hare, Robert, D. 2003: *Without Conscience: The Disturbing World of the Psychopaths Among Us*. New York, NY: The Guilford Press.
- Hassabis, D. K. D., S. D. Vann, and Eleanor A. Maguire 2007: 'Patients with Hippocampal Amnesia Cannot Imagine New Experiences'. *Proceedings of the National Academy of Sciences*, 104, pp. 1726–31.
- Hauser, M. D. 2006: 'The Liver and the Moral Organ'. *Social Cognitive and Affective Neuroscience*, 1, pp. 214–20.
- Heims, H. C., H. D. Critchley, R. Dolan, C. J. Mathiasand, and L. Cipolotti 2004: 'Social and Motivational Functioning is not Critically Dependent on Feedback of Autonomic Responses: Neuropsychological Evidence from Patients with Pure Autonomic Failure'. *Neuropsychologia*, 42(14), pp. 1979–88.
- Kapur, S., C. Jones, G. M. Brown, S. Houle, and E. Tulving 1995: 'Functional Role of the Prefrontal Cortex in Retrieval of Memories: A PET Study'. *Neuroreport*, 14, pp. 1880–4.
- Keane, M. M., J. D. Gabrieli, H. Mapstone, K. A. Johnson, and S. Corkin 1995: 'Double Dissociation of Memory Capacities after Bilateral Occipital-lobe or Medial Temporal-lobe Lesions'. *Brain*, 118, pp. 1129–48.
- Kennett, J. and C. Fine 2008a: 'Internalism and the Evidence from Psychopaths and "Acquired Sociopaths"'. In Sinnott-Armstrong 2008, pp. 173–90.
- 2008b: 'Could there be an Empirical Test of Internalism?' In Sinnott-Armstrong 2008, pp. 217–26.
- 2009: 'Would the Real Moral Judgment please stand up? The Implications of Social Intuitionist Models of Cognition for Meta-ethics and Moral Psychology'. *Ethical Theory and Moral Practice*, 12, pp. 77–96.

- Klein, S., J. Loftus, and J. F. Kihlstrom 2002: 'Memory and Temporal Experience: the Effects of Episodic Memory Loss on An Amnesic Patient's Ability to Remember the Past and Imagine the Future'. *Social Cognition*, 20, pp. 353–79.
- Knight, R. T. 1999: 'Prefrontal Cortex Regulates Inhibition and Excitation in Distributed Neural Networks'. *Acta Psychiatrica Scandinavica*, 101, pp. 159–78.
- Kochanska, G., D. L. Padavich, and A. L. Koenig 1996: 'Children's Narratives About Hypothetical Moral Dilemmas and Objective Measures of Their Conscience: Mutual Relations and Socialization Antecedents'. *Child Development*, 67, pp. 1420–36.
- Kochanska, G., K. Murray, and K. C. Coy 1997: 'Inhibitory Control as a Contributor to Conscience in Childhood: From Toddler to Early School Age'. *Child Development*, 68, pp. 263–77.
- Koenigs, M., L. Young, R. Adolphs, D. Tranel, F. Cushman, M. Hauser, and A. R. Damasio 2007: 'Damage to the Prefrontal Cortex Increases Utilitarian Moral Judgements'. *Nature*, 446, pp. 908–11.
- Kolak, D. and R. Martin 1991: *Self and Identity: Contemporary Philosophical Issues*. New York, NY: Macmillan.
- Korsgaard, C. 1988: 'Personal Identity and the Unity of Agency: A Kantian Response to Parfit'. *Philosophy and Public Affairs*, 18, pp. 101–32.
- 2002: 'Internalism and the Sources of Normativity'. In Pauer-Studer 2002, pp. 49–69.
- Levin, Harvey S., Howard M. Eisenberg, and Arthur Lester Benton (eds) 1991: *Frontal Lobe Function and Dysfunction*. New York: Oxford University Press.
- Levine, B., D. Dawson, S. E. Black, and D. T. Stuss 1999: 'Ventral Frontal Contribution to Self-regulation: Convergence of Episodic Memory and Inhibition'. *Neurocase*, 5, pp. 263–75.
- Levine, B., G. T. Turner, D. J. Tisserand, S. J. Graham, S. Hevenor, and A. R. McIntosh 2004: 'The Functional Neuroanatomy of Episodic and Semantic Autobiographical Remembering: A Prospective Study'. *Journal of Cognitive Neuroscience*, 16, pp. 1633–46.
- Maia, T. V. and J. L. McClelland 2004: 'A Reexamination of the Evidence for the Somatic Marker Hypothesis: What Participants Really Know in the Iowa Gambling Task'. *Proceedings of the National Academy for Science USA*, 101(45), pp. 16,075–80.

- 2005: 'The Somatic Marker Hypothesis: Still Many Questions but no Answers'. *Trends in Cognitive Science*, 9, pp. 162–4.
- Miller, E. and J. D. Cohen 2001: 'An Integrative Theory of Prefrontal Cortex Function'. *Annual Review of Neuroscience*, 24, pp. 167–202.
- Moll, J., R. de Oliveira-Souza, P. J. Eslinger, I. Bramati, J. Mourao-Miranda, A. Andreiueolo, and L. Pessoa 2002: 'The Neural Correlates of Moral Sensitivity: a Functional Magnetic Resonance Imaging Investigation of Basic and Moral Emotions'. *Journal of Neuroscience*, 22(7), pp. 2730–6.
- Nichols, S. 2002: 'How Psychopaths Threaten Moral Rationalism: Is It Irrational to be Immoral?' *Monist*, 85, pp. 285–304.
- 2004: *Sentimental Rules: On the Natural Foundations of Moral Judgment*. Oxford: Oxford University Press.
- North, N. T. and R. E. O'Carroll 2001: 'Decision Making in Patients with Spinal Cord Damage: Afferent Feedback and the Somatic Marker Hypothesis'. *Neuropsychologia*, 39, pp. 521–4.
- Prinz, J. 2006: 'The Emotional Basis of Moral Judgments'. *Philosophical Explorations*, 9, pp. 29–44.
- Pauer-Studer, Herlinde (ed.) 2002: *Constructions of Practical Reason: Interviews on Moral and Political Philosophy*. Stanford: Stanford University Press.
- Pylyshyn, Z. 1999: 'Is Vision Continuous with Cognition? The Case for Cognitive Impenetrability'. *Behavioural and Brain Sciences*, 22, pp. 341–423.
- Roskies, A. 2003: 'Are Ethical Judgements Intrinsically Motivational? Lessons from "Acquired Sociopathy"'. *Philosophical Psychology*, 16, pp. 51–66.
- 2008: 'Internalism and the Evidence from Pathology'. In Sinnott-Armstrong 2008, pp. 191–206.
- Saltzstein, D. Herbert and Tziporah Kasachkoff 2004: 'Haidt's Moral Intuitionist Theory: A Psychological and Philosophical Critique'. *Review of General Psychology*, 8(4), pp. 273–82.
- Saver, J. L. and A. R. Damasio 1991: 'Preserved Access and Processing of Social Knowledge in a Patient with Acquired Sociopathy due to Ventromedial Frontal Damage'. *Neuropsychologia*, 29(12), pp. 1241–9.
- Schacter, D. L., K. Norman, and W. Koutsaal 1998: 'The Cognitive Neuroscience of Constructive Memory'. *Annual Review of Psychology*, 49, pp. 289–318.

- Schacter, D. L., D. R. Addis, and R. L. Buckner 2007: 'Remembering the Past to Imagine the Future: The Prospective Brain'. *Nature Reviews Neuroscience*, 8, pp. 657–61.
- 2008: 'Episodic Simulation of Future Events: Concepts, Data, and Applications'. *Annals of the New York Academy of Sciences*, 1124, pp. 39–60.
- Sinnott-Armstrong, W. (ed.) 2008: *Moral Psychology, Volume 3: The Neuroscience of Morality: Emotion, Brain Disorders, and Development*. Cambridge, MA: MIT Press.
- Smith, M. 1994: *The Moral Problem*. Oxford: Blackwell.
- Steinvorth, S., S. Corkin, and E. Halgren 2006: 'Ecphory of Autobiographical Memories: an MRI Study of Recent and Remote Memory Retrieval'. *Neuroimage*, 30, pp. 285–98.
- Strauss, J. and G. R. Goethals (eds) 1991: *The Self: Interdisciplinary Approaches*. New York: Springer-Verlag.
- Stuss, D. T. 1991: 'Self, Awareness, and the Frontal Lobes: A Neuropsychological Perspective'. In Strauss and Goethals 1991, pp. 255–78.
- Suddendorf, T. and M. Corballis 1997: 'Mental Time Travel and the Evolution of the Human Mind'. *Genetic, Social, and General Psychology Monographs*, 123, pp. 133–67.
- 2008: 'The Evolution of Foresight: What is Mental Time Travel and is it Unique to Humans?' *Behavioural and Brain Sciences*, 30, pp. 299–313.
- Suddendorf, T. and J. Busby 2005: 'Making Decisions with the Future in Mind: Developmental and Comparative Identification of Mental Time Travel'. *Learning & Motivation*, 36, pp. 110–25.
- Tomb, I., M. Hauser, P. Deldin, and A. Caramazza 2002: 'Do Somatic Markers Mediate Decisions on the Gambling Task?' *Nature Neuroscience*, 5, pp. 1103–4, (author reply p. 1104).
- Tranel, D. and A. R. Damasio 1988a: 'Knowledge Without Awareness: An Autonomic Index of Facial Recognition in Prosopagnosics'. *Science*, 228, pp. 1453–4.
- 1988b: 'Non-conscious Face Recognition in Patients with Face Agnosia'. *Behavioural Brain Research*, 30, pp. 235–49.
- Tranel, D., H. Damasio, and A. R. Damasio 1995: 'Double Dissociation Between Overt and Covert Recognition'. *Journal of Cognitive Neuroscience*, 7, pp. 425–32.
- Tulving, E. 1985: 'Memory and Consciousness'. *Canadian Psychologist*, 26, pp. 1–12.

- 2002: 'Episodic Memory: From Mind to Brain'. *Annual Review of Psychology*, 53, pp. 1–25.
- Velleman, J. D. 1991: 'Well-Being and Time', in his 2000, pp. 56–84. Originally published in *Pacific Philosophical Quarterly*, 72, pp. 48–77.
- 1997: 'Deciding How to Decide'. In G. Cullity and B. Gaut 1997, pp. 29–52.
- 2000: *The Possibility of Practical Reason*. Oxford: Clarendon Press.
- Wallace, R. J. 1999: 'Moral Cognitivism and Motivation'. *Philosophical Review*, 108, pp. 161–219.
- Waltz, J., B. Knowlton, K. Holyoak, K. Boone, F. Mishkin, M. DeMenezes Santos, C. Thomas, and B. Miller 1999: 'A System for Relational Reasoning in the Prefrontal Cortex'. *Psychological Science*, 10, pp. 119–25.
- Weiskrantz, L., E. K. Warrington, M. D. Sanders, and J. Marshall 1974: 'Visual Capacity in the Hemianopic Field Following a Restricted Occipital Ablation'. *Brain*, 97, pp. 709–28.
- Wheatley, T. and J. Haidt 2005: 'Hypnotically Induced Disgust Makes Moral Judgments More Severe'. *Psychological Science*, 16, pp. 780–4.
- Wheeler, M. A., D. Stuss, and E. Tulving 1997: 'Toward a Theory of Episodic Memory: The Frontal Lobes and Autonoetic Consciousness'. *Psychological Bulletin*, 121, pp. 331–54.
- Wood, J. and J. Grafman 2003: 'Human Prefrontal Cortex'. *Nature Reviews Neuroscience*, 4, pp. 139–47.
- Zentall, T. R. 2006: 'Mental Time Travel in Animals: A Challenging Question'. *Behavioural Processes*, 72, pp. 173–83.