The Neural Mechanisms of Moral Cognition: A Multiple-Aspect Approach to Moral Judgment and Decision-Making

WILLIAM D. CASEBEER[1]
Department of Philosophy
HQ USAFA/DFPY
2354 Fairchild Drive
Suite 1A10
US Air Force Academy, CO 80840
U.S.A.
E-mail: William.casebeer@usafa.af.mil

PATRICIA S. CHURCHLAND
Philosophy Department 0119
University of California at San Diego
La Jolla, CA 92093
U.S.A.
E-mail: pschurchland@ucsd.edu

Abstract. We critically review the mushrooming literature addressing the neural mechanisms of moral cognition (NMMC), reaching the following broad conclusions: (1) research mainly focuses on three inter-related categories: the moral emotions, moral social cognition, and abstract moral reasoning. (2) Research varies in terms of whether it deploys *ecologically valid* or *experimentally simplified* conceptions of moral cognition. The more ecologically valid the experimental regime, the broader the brain areas involved. (3) Much of the research depends on simplifying assumptions about the domain of moral reasoning that are motivated by the need to make experimental progress. This is a valuable beginning, but as more is understood about the neural mechanisms of decision-making, more realistic conceptions will need to replace the simplified conceptions. (4) The neural correlates of real-life moral cognition are unlikely to consist in anything remotely like a "moral module" or a "morality center." Moral representations, deliberations and decisions are probably highly distributed and not confined to any *particular* brain sub-system. Discovering the basic neural principles governing planning, judgment and decision-making will require vastly more basic research in neuroscience, but correlating activity in certain brain regions with well-defined psychological conditions helps guide neural level research. Progress on social phenomena will also require theoretical innovation in understanding the brain's distinctly *biological* form of computation that is anchored by emotions, needs, drives, and the instinct for survival.

Key words: affect, amygdala, ethics, evolution, hippocampus, moral cognition, moral psychology, naturalization, neuroscience, neurobiology, prefrontal cortex, theory of mind

Introduction

The neurobiology of moral cognition is a justifiably hot topic. In the past two years, discussions focusing on it have made the front pages of such publications as *The New York Times, Time* and *Newsweek*. So far, however, there has been no concerted effort to integrate the findings of the various experiments to provide a broader and more comprehensive picture.[2] Here, we attempt such integration, arguing that this perspective will teach us several things about both the assumptions that underlay various research programs and the nature of a comprehensive theory of moral cognition.

We say "justifiably hot," rather than merely faddish, because of the critical importance of good moral judgment to civilized life. Exploring the neural roots of moral cognition promises to give us more informed theories of moral reasoning, help us identify moral pathologies so they might be alleviated or avoided, and allow us to improve our moral education and training.[3] The experimental literature can shed light on traditional issues in moral psychology, and is critical for the enterprise of naturalizing ethics – of showing how our ability to grasp norms, reason about them, and act on them is a brain-based ability that can be explored using scientific methods.

After preliminary definitions and some brief discussion of moral theories and their implicit assumptions about the nature of the psychology of morality, we canvass the empirical work targeting the moral emotions, moral social cognition, and abstract moral reasoning, before reaching general conclusions about the form and nature of the neural mechanisms of moral cognition.

Defining moral cognition

What is moral cognition? What is a moral judgment? In any inquiry, delimiting the domain is critical (for example, if you are interested in studying the neural correlates of *face recognition*, you need to have a working conception of just what constitutes a *face*). In the moral realm, this problem is magnified many times over, as there is disagreement even among professional ethicists regarding what we should properly call a moral judgment or an act of moral cognition. We need to cast our net widely enough such that we don't eliminate what are in fact aspects of moral reasoning, but narrowly enough such that we identify a bona fide and manageable area of study.

Two items to note: first, on any theory of what morality consists in, moral reasoning will involve a series of cognitive acts that issue in a conclusion (either implicit or explicit) about what one *ought* to *do* or *think*. As moral philosophers would put it, moral reasoning is probably a species of practical reasoning – reasoning about what we should do or think now, such as whether

to negotiate with terrorists, not necessarily about what others have done nor about strictly empirical matters, such as whether there is water on Mars. Moral cognition thus deals with cognition about *norms* rather than cognition about *facts*. Nevertheless, since good reasoning about norms is knowledge-dependent, the difference is one of degree rather than kind. Broadly speaking, then, moral reasoning deals with *cognitive acts and judgments associated with norms, or with facts as they relate to norms*. Note that the term "judgment" here should not be used in a linguistically chauvinistic manner. Social animals such as wolves and baboons do show understanding of conspecific norms, as well as of the groups' local conventions, pecking order, division of labor, and even who has what kind of knowledge (Brothers 1997).

Second, how widely you cast your moral cognitive net varies according to your background substantive theory of morality, in much the same way that your face recognition net's width would vary according to your background theory about faces. This may seem to present us with a vicious circularity: we need to have a normative theory of morality in place before we can identify the neural correlates of moral cognition; but on the other hand, one advantage of identifying the neural correlates of cognition is that it may allow us to eliminate certain moral theories as being psychologically and neurobiologically unrealistic. However, this circularity (appearances to the contrary) is *not* a vicious one: theories and their domains *co-evolve*, informing one another, in many areas of the sciences. To continue with our face recognition example, experiments accomplished by Fantz in 1961 demonstrated that babies as young as two months have a preference to attend to faces.[4] Work by Johnson and Morton in 1991 showed that newborns even attend preferentially to just a triangular arrangement of three blobs. Studies such as these used objects with face-like features to *simulate* faces; some critics argued that these were not enough like *actual faces* to justify the contention that these babies have a preference for faces as such. Further research has used stimuli that are more like traditional faces (although puppets used while studying newborns still sometimes have only a "family resemblance" relation to actual adult physiognomies), but researchers have reduced the demands they make upon face stimuli as they have recognized the value of being able to explore face recognition capacities using alternate stimuli regimens.

This give-and-take process between theory and domain may result in a domain that is "fuzzy" rather than sharply delineated. For example, if someone were to identify the neural correlates of musical cognition, we might expect different brain resources to be recruited depending upon whether the music is emotionally resonant, or depending upon whether the subjects listen to Ludwig Beethoven's "Fifth Symphony" or John Cage's "Four Minutes Thirty-Three Seconds" of silence. This does not mean, though,

that the domain of musical cognition is not worthy of study nor that it is "unscientific," especially if at least some reliably recruited subsystems are involved. Categories that are *strikingly* ad hoc (for example, the neural correlates of basketball playing), may best be analyzed by breaking them down into their constituent "natural-cognitive-kind" components, of course; and it may very well turn out that empirical investigation of any domain might actually *eliminate* that domain as a contender for the identification of uniquely involved brain regions. In the case of moral cognition, the empirical work has illuminated some reliably involved brain systems, but as we will argue later, it is highly unlikely, given the data, that we shall find anything like a well-defined "moral reasoning module" in the mind/brain.

Our initial take on the domain of moral cognition, then, can and should be informed by a background moral theory (in our case, a neo-Aristotelian virtue theory, according to which moral concerns relate to what we have to *think* and *do* so as to *function well* as human beings). Related pragmatic considerations are useful too: what do professionals say about the moral domain and what are the points of contention? What has the extant research considered the proper domain of moral cognition? What do intermediate and high level theories (such as "pure" moral psychology) have to say about the domain? In the review that follows, we use all these factors to identify the literature to summarize.

A Brief tour of major moral theories and their associated psychologies

It will be useful to keep in mind the three major extant normative moral theories and the assumptions they make about human psychology. The first theory is the venerable and ancient virtue theory. First articulated by Plato and Aristotle over 2,000 years ago, it maintains that moral questions are questions regarding what kind of people we must be so as to function as best we can, identifying certain character traits that are necessary for us to function properly. Moral judgments are judgments about what traits we must cultivate so as to live maximally functional lives. Moral cognition (in its broadest sense) involves a far-reaching understanding that relates to our well-being (in the broadest sense). The emphasis in this theory is on "knowing how" to be virtuous, not merely on "knowing that" a certain act would be praiseworthy. It focuses as much on skill, habit and praxis as on theory and abstraction. Virtue theorists pay close attention to the role of the emotions in ethical affairs, assigning it an important role in moral education and the moral life. Virtue theory is generally thought to have the richest moral psychology of the major moral theories, as reason, appetite, emotion, and affect all play vital roles in the mental life of an effective moral agent.

Utilitarianism, Jeremy Bentham and John Stuart Mill's seventeenth and eighteenth century ethical theory, has as its guiding principle the idea that states of pleasure and pain (in a rich, sophisticated sense, not just in a "pig-like" sense) are the obvious "goods" and "bads" in the universe. We are obliged, then, to take those actions or follow those rules that, if taken or followed, would produce the greatest amount of happiness for the greatest number of sentient beings (where sentience is an indicator of your ability to experience pleasure and pain). A utilitarian's moral psychology would also be fairly rich; a good utilitarian reasoner would be especially effective at forecasting the consequences of actions and at identifying particular actions that would be conducive to the greatest happiness for the greatest number, and would be motivated to take such actions.

The third major moral theory is deontology, from the Greek *deon* for "duty" (literally: the study of our duties). This theory identifies morality with doing one's duty, and in its most famous version (articulated by Prussian philosopher Immanuel Kant in the eighteenth century), we use the dictates of pure reason to discover what duties we owe to other free and reasonable creatures. The principal test we use to discover our duty is the Categorical Imperative, which tells us (in two of its several formulations) that we should act only on those maxims that we can will to become a *universal law*, and that we ought not to treat others *merely as means* but always also as ends worthy of respect. Morality consists in doing your duty for the sake of duty alone, and not because it will gratify your nerve endings or because your emotions move you to take action. What kind of cognitive faculties are posited by the Kantian system? The ability to reason "purely," for one. Exactly what this capacity consists in is difficult to determine. At the very least, however, it involves formulating language-like maxims that are then checked for consistency. If emotion or affect tags or marks the maxims and the associated logical processing that is accomplished over them, then the reason is not pure – either we will fail to respect reason as such owing to bad advice from our limbic system, or our maxim will become tainted with the inappropriate motivation even if we 'do the right thing' (remember, we should respect reason – do our duty – for duty's sake alone).

To a first approximation, neuroscientific results suggest that the moral psychology of humans and other social animals such as chimpanzees, baboons, ravens, and wolves, are more consilient with the assumptions of virtue theory than with other theories. To justify this hypothesis, we draw on a selected sample of results that reveal strong links between emotion, social cognition, explicit learning (declarative memory), and implicit learning (skills, habits, conditioning, priming). In addition, we suggest that considerations from evolutionary biology, developmental psychology, and develop-
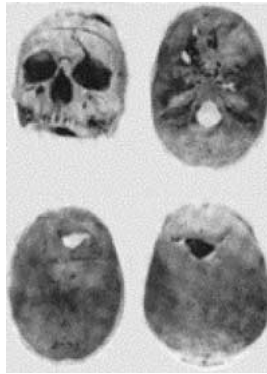
Figure 1. Gage's skull. Courtesy of the Warren Museum.

mental neuroscience, also favor this hypothesis (see Casebeer, forthcoming; P. Churchland 1998; P.S. Churchland 2002; Damasio, forthcoming).

Research on the neurobiology of the moral emotions in moral reasoning

A famous accidental experiment: Phineas Gage

One of the earliest and most famous "experiments" involves the accidental injury of railway foreman Phineas Gage late last century. On September 13, 1848, an accident sent a dynamite tamping iron through (among other structures) Gage's prefrontal cortex in both hemispheres (see Figure 1). It is reasonable to infer that the ventral and medial areas of Gage's prefrontal lobes were all but destroyed. Following the accident, Gage was a changed man; he became unreliable at work and eventually became a homeless drifter and alcoholic.[5] His motor control and sensory perception, however, were normal, so far as the evidence shows. While it is difficult to know exactly what parts of Gage's brain were damaged, we do know that his social behavior changed dramatically following the accident. Later research revealed this was most likely due to bilateral damage to his prefrontal cortex caused by the passage of the rod. The Gage case is important because it provoked study on the role of the frontal structures in temperament and self-control and suggested a link between reasoning and the emotions.[6]

Prefrontal cortex (PFC)

In mammals and reptiles, the brain's regulatory core is situated in the brain-stem/limbic axis (see Figure 2). Basic regulatory functions that depend on
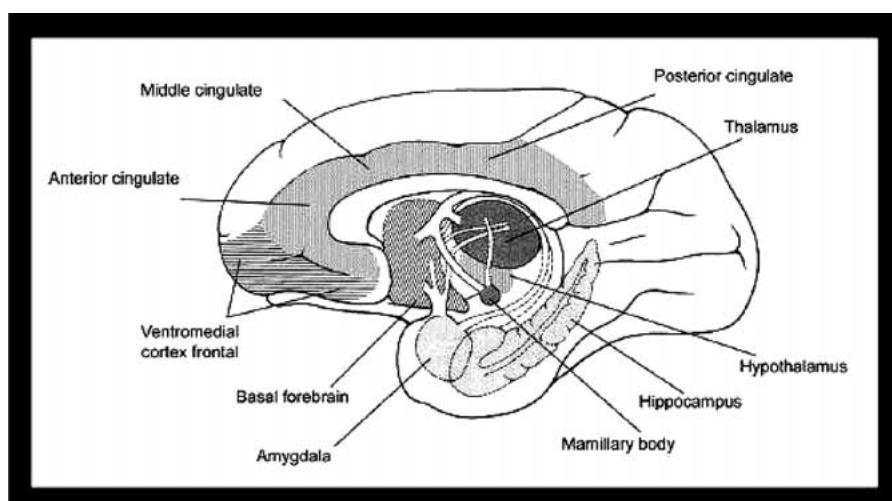
Figure 2. A drawing showing some of the limbic structures.

brainstem circuitry include breathing, blood pressure, arousal, thermoregulation, shifts in behavioral state from wakeness to sleep to dreaming, integration of signals from the all-pervasive interoceptive system that carries signals about a host of features of the inner body, and coordination of inner drives (for food, sex, oxygen, etc.) with perceptions (e.g., flee now – do not feed). The emotions associated with anoxia, hunger, thirst, sexual desire, and fear of being killed are very powerful. The evolution of larger brains capable of adaptive social behavior involved an expansion of frontal cortex, with preservation of the regulatory platform in the brainstem/limbic axis and its innervation of frontal structures. The cortical expansion permitted more sophisticated predictive capacities that relied upon more complex motor planning and decision-making, more complex attentional operations, and more complex relationships between the range of emotions, drives, and behavior.

Humans share most of their genes with other vertebrates, and the brain is organized in essentially the same way in all mammals. Although it is not impossible that moral behavior as seen in humans is produced by a totally new brain structure, such novelty looks improbable given what is known about brain organization and development. On the contrary, moral behavior in humans appears closely linked to social behavior in other species, and differs mainly in level of sophistication, some of which is undoubtedly language-dependent. From an evolutionary perspective, the survival of the animal depends on its maintaining its inner milieu within a very narrow range of values. Consequently, however fancy the capacities for predicting, defer-

ring gratification, recognizing and behaving in accordance with social rank and order, the framework holding it all together is the emotion/drive/body-preserving circuitry in the brainstem/limbic axis.

Prefrontal cortex (PFC) is defined at that part of frontal cortex that lies anterior and medial to the motor and premotor cortex (see Figure 3). In primates, the territory is vast, and is far less well understood both anatomically and physiologically than the sensory systems in the more posterior regions of the brain. Experimentally, prefrontal cortex has been much more difficult to explore than sensory systems, largely because its neurons are many synapses distant both from the receptor surfaces and from the motoneuron-muscle interface. Given PFC input from brainstem and subcortical structures, it is not surprising that the specific neuronal response patterns, even when found, may be difficult to interpret.

Nevertheless, some important neuron-level results have been obtained. Single neuron studies in monkeys have revealed a specific region of the dorsolateral cortex that respond preferentially during a delay period to holding specific kinds of information, such as the precise spatial location of a target (See Fuster 1995; Goldman-Rakic 1987). The correlative lesion and imaging studies confirm the importance of this area for holding information on-line in working memory. This is one of the first results at the neuronal level that ties a specific function (holding spatial information during a delay) to response patterns of specific neurons (those that fire only in the absence of the stimulus and that show preference for a specific position in space).

A second important set of results from prefrontal cortex concerns individual neurons that respond when either the monkey *makes* a specific movement, such as grasping with thumb and index finger, or when it sees another make precisely the same movement (Rizzolatti, Fogassi and Gallese 2001). The behavior of these "mirror neurons" suggests that when seeing the other make the movement, the premotor cortex generates incipient motor commands to match the movement. It is possible that these signals can be detected as intentions, albeit off-line intentions, which are used to interpret what is seen (e.g., "he intends to share food"). Related results now show other neuron populations tuned to both executing a specific action and *hearing* another perform that very action.

The mirror-neuron research helps begin to explain how even newborns can imitate actions such as grimacing and sticking out the tongue. It also motivates the hypothesis that *inner simulation* of other behavior is an important ingredient in development of a representational network known as a "theory of mind" (ToM). Just as brains represent the causal and categorial structure of the world (e.g., "nettles can sting"), so they represent the causal and categorial structure of conspecific behavior (e.g., "he will get angry if I refuse to groom
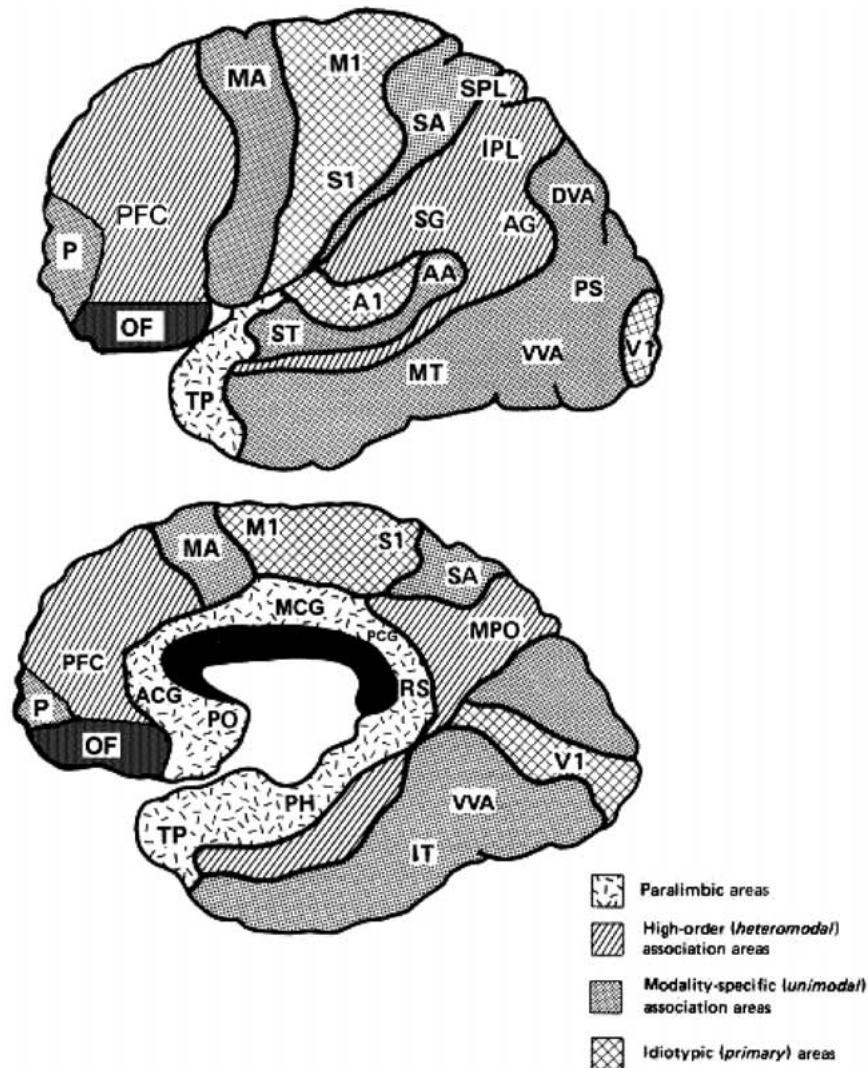
Figure 3. A schematic characterization of cortical areas in relation to distance from the sensory or motor periphery. Upper: lateral aspect. Lower: medial aspect. PFC: prefrontal cortex; OFC, orbital-frontal cortex; F, frontal pole; ACG, anterior cingulate cortex; MCG, middle cingulate cortex; PCG, posterior cingulate cortex; PH, parahippocampal area; M1, primary motor area; MA; supplementary motor areas; S1, primary somatosensory area; SA; associative somatosensory area; SPL, superior parietal lobe; IPL, inferior parietal lobe; SG, supramarginal gyrus; AG, angular gyrus; A1, primary auditory area; TP, temporal pole; ST, superior temporal gyrus; IT, inferior temporal; MT, medial temporal; MPO, medial parietal-occipital; VVA, ventral visual area; DVA, dorsal visual area, V1, primary visual cortex; PO, preoptic area; CG cingulate gyrus. Adapted from Cytowic (1996).

him"). A ToM network is probably what enables us to make predictions regarding what others are intending or will feel if I do *this*. It is probably what enables a chimpanzee, for example, to know whether a high-ranking male across the table is able to see the food she is grabbing or whether the food is occluded from his view by a box and can be safely taken (Call 2001).

Synaptogenesis and neuronal mylenation in PFC occurs significantly later in development than in sensory and motor systems. This is consistent with the behavioral data on the development of emotional and behavioral maturity in children. For example, though infants reliably imitate sticking out the tongue, adults do not. Three-year olds explain and predict what other humans do mainly by reference to desires and perceptions, but are not yet in command of the notion of belief. Three-year olds can use counterfactuals about desires and can easily answer such questions as "if Billy wanted a cookie and I gave him a pencil would he be happy?" Whereas children at age four typically can correctly say that their friend will falsely *believe* that there are candies in a box, children at age three typically cannot. The emergence in children of a full-blown ToM corresponds with the ability to lie,[7] suggesting that they understand what will be falsely believed by the other person and what that person will accordingly do. To a first approximation, improvements in capacities for impulse-control, self-regulation, and ToM sophistication appear to track synaptogenesis and mylenation in PFC.

Lesion studies of humans with focal brain damage as well as experimental focal lesions on monkeys have demonstrated a relationship between prefrontal tissue and planning, decision-making, emotion, attention, memory for spatial and temporal patterns, and recognition of a mismatch between intent and execution. This is a large set of complex functions, and so far, the precise nature of the relationship between these functions and the specific nature of the contribution made by PFC remains unclear. Within the last decade, the use of imaging techniques such as functional magnetic resonance imaging (fMRI) has encouraged new probes into prefrontal function in humans and monkeys, though the spatial and temporal resolution of imaging techniques is still poor. Since there are about a million synapses in a cubic millimeter of cortical tissue, a technique with spatial resolution in the millimeter range is, though useful, still very crude. In particular, it will be difficult to determine the precise functions and their dependencies unless we have techniques with finer resolution than what is provided by existing imaging technology. Likewise, it will be difficult to determine *network* properties until we have high-temporal resolution techniques for accessing thousands of neuronal responses simultaneously. Nevertheless, it should be noted that an important contribution of imaging studies consists in guiding higher-
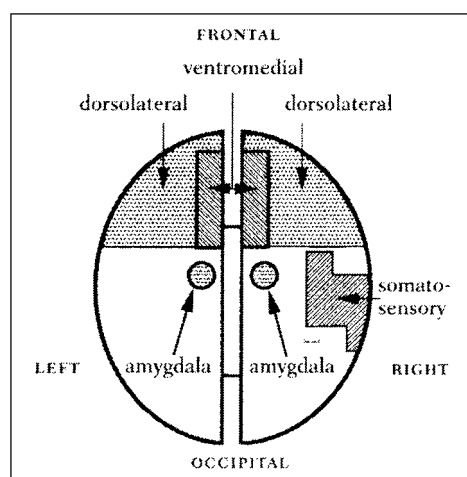
Figure 4. A schematic characterization of location of ventromedial cortex, the prefrontal region, and its relation to somatosensory corteces and the insula. Courtesy of Hanna Damasio.

resolution techniques to explore specific areas and indicating the distribution of areas showing heightened activity in precisely specified conditions.

Despite the difficulties in exploring prefrontal regions, the link between moral decision-making, social cognition, and the emotions is beginning to emerge. Converging results from lesions studies and imaging studies indicate that ventromedial (VM) prefrontal damage is consistently associated with impairments in practical decision-making, moral decision-making, emotion, and feeling (see for example, Saver and Damasio (1991), Damasio (1994), Bechara et al. (1994), and Anderson et al. (1999), Figure 4) Patients with focal VM lesions show abnormally flat responses when shown emotional pictures, and they also show poor results on tasks where they need to rely on feelings to guide complex self-directed choices. These findings are consistent with what is known about the neuronal pathways connecting limbic and brainstem structures to VM (see Figure 4).

Magneto Encephalogram (MEG) studies of youngsters with conduct and self-control disorders (Bauer and Hesselbrock 2002) show a similar result. The finding has also been confirmed with fMRI work involving multiple normal subjects (such as Moll et al.'s 2002a, b studies which detected VM, as well as superior temporal sulcus, recruitment when subjects viewed scenes evocative of moral emotions) and by Positron Emission Tomography (PET) studies of multiple normal humans (see Houdé et al. 2001).

Within the ventral prefrontal cortex, orbitofrontal cortex (OFC) is the large cortical region lying just above the orbits of the eye. Evidence suggests it plays a critical role in *cuing* morally appropriate behavior in adulthood and

in *acquiring* moral knowledge in childhood. For example, Anderson et al. (1999) found that while patients with adult-onset and childhood-onset OFC damage displayed similar abnormal social and moral behavior, their scores on standardized tests of moral reasoning were markedly different. Those who suffered from *early* onset damage performed remarkably poorly, exhibiting overly egoistic moral reasoning more appropriate for a ten year old, whereas adult onset subjects performed normally on the standardized tests. Connections between the reinforcement learning circuitry – in the interoceptive system, the amygdala, the hippocampal structures, and OFC and mediated by the dopamine system – seem to be necessary for acquiring moral skills, social perceptions, and an understanding of social expectations, whether or not the knowledge can be articulated. To a first approximation, if the feelings incident upon punishment and approval in its various forms cannot be associated with specific social events, moral and life-skill learning are impaired (Baxter and Murray 2002). Simple conditioning, for example, associating a tone with a pain, is still possible, but with OFC damage, the complex cognitive processes involved in social learning and social cognition are disconnected from feelings that arise when social disutilities are visited upon an agent. It is also worth noting that OFC shows a relative decline in activity during REM sleep, which is consistent with the decreased social inhibition and lack of insight that typifies dreams. Limbic structures, by contrast, are relatively active during REM, which again is consistent with the frequency of strong emotional components in dreams (see also Kahn et al. 2002).

To round out the connectivity story, it should be noted that the PFC is reciprocally connected not only to subcortical structures but to sensory corteces beyond their primary areas; i.e. extrastriate visual cortex, S2 (somatosensory cortex), and A2 (auditory cortex). These pathways presumably provide perceptual signals relevant to guiding PFC operations. Although textbooks commonly assign PFC an "integrative and executive" role, this description will continue to be almost vacuous until much more of the basic neurobiology is understood. That is, in PFC we need anatomical identification of neuron types, the physiology of specific neuron types, the principles governing plasticity in the region, and more detail about the precise nature of the connectivity patterns and what causes what.

In the early stages of science, we often have to make do with rough ideas. We do have rough ideas, based on behavioral studies under a range of conditions and with a range of animals, that *integrative* functions involves drawing on a broad range of signals to extract a new and more abstract representation, and that *executive* functions have something to do with choice and motor behavior that draw on skills and habits, as well as current perceptions. This is vague, but it can be improved, revised and reconstructed as science proceeds.

Since PFC functions are critically dependent on signals from limbic structures, we briefly profile some aspects of limbic function.

Limbic structures, emotions and drives

The limbic structures are a highly interconnected set of subcortical regions that include the hippocampus, amygdala, basal forebrain and hypothalamus. The only cortical structure included is the cingulate cortex – that part of the medial cortex that forms a band bordering the cerebral commissures. Activity in all these structures is modulated by the main four neurotransmitter systems originating in distinct brainstem nuclei and identified in terms of their specific transmitter: dopamine, serotonin, noradrenalin and acetylcholine. Changes in these systems can have powerful effects on cognition, consciousness, anxiety levels, aggressiveness, sexual drive, moods and on the emotions. For example, as we fall asleep, the activity of the noradrenalin and serotonin systems tapers off, and at their lowest level, we are in deep sleep. As the REM period approaches, acetylcholine levels increase and at their highest levels, we are in full dreaming state. Lowering levels of serotonin in monkeys changes social rankings, and results in self-defeating aggression and reckless behavior. Males with a mutation in a gene that results in low serotonin levels (MAOA gene on the X chromosome) and who are maltreated in childhood are about three times as likely to have conduct disorders and display irrational violence than those who have the normal gene, and about twice as likely as those who have the gene but had a normal childhood (Caspi et al. 2002).

The amygdala

The amygdala is a multi-component structure, highly connected to prefrontal cortex, and to other limbic structures, including the hippocampus and the basal forebrain. One specific component (the lateral nucleus) has been identified as crucial for aversive-conditioning, for negative feelings such as fear, and for recognizing a situation as fearful and a face as showing fear. Other regions of the amygdala form part of the complex reward circuitry involving the positive emotions (Hamann et al. 2002). These regions are also implicated in a range of addictions. A plausible hypothesis is that among other things, the amygdaloid complex modulates storage of emotionally significant or arousing events. Roughly, perceptions involving events of great importance to the organism's survival provoke specific emotions (anger, fear, delight, disgust) and with amygdala activity, are more likely to be permanently stored than perceptions involving emotionally neutral events. The amygdala's reciprocal connections to hormonal systems, including glucocorticoid (stress hormone) system and the opiod systems also lend credence to the hypothesis. Considerable evidence links the amygdala with various forms of implicit or

nonconscious memory (including conditioning), and its connections to the hippocampus indicate it also serves in explicit memory.[8]

Adolphs et al. (1998) report that the amygdala is critical for aiding retrieval of socially relevant knowledge about facial appearance. Three subjects with complete bilateral amygdala damage were asked to judge the approachability and trustworthiness of unfamiliar people. All three subjects judged unfamiliar people to be more approachable and more trustworthy than did controls. Interestingly, the subjects were able to make normal judgments when the descriptions were *verbal* rather than *visual*. This indicates the amygdala is particularly critical for making appropriate moral judgments about visual stimuli (especially faces), while the processing requirements for words and sentences tap a broader range of information-laden brain structures in frontal cortex.

Recent work by Weinberger et al. (2002) indicates that proper amygdala function as it affects general temperament might be regulated by one critical gene sequence. Using fMRI on twenty-eight subjects, those whose DNA contained a shorter than normal length of a critical promoter gene for a protein that assists the distribution of serotonin, showed heightened activity in their right amygdala while processing a task related to matching angry and frightened faces. Bluntly: subjects with the short allele are more prone to anxiety and to feeling threatened even in non-threatening situations, which might lead to a tendency towards dysfunctional sociability.

The cingulate cortex (CC)

The cingulate cortex has a number of subregions that appear to make somewhat different contributions to normal brain function. Regulation of selective attention, regulation of motivation, and detection of mismatch between intention and execution are associated with anterior regions (ACC) (Figure 5). Work in both humans and monkeys indicates that the ACC conveys reinforcing stimuli to other parts of the cortex, and plays a key role in addictive and obsessive-compulsive disorders. Montague et al. (2002) indicate that leveraging rostral ACC activation (along with the nucleus accumbens, the caudate nucleus, and VM/orbitofrontal cortex) is needed for cooperative behavior among subjects playing a version of the "Prisoner's Dilemma" (as the popular press said *brains want to cooperate because it feels good to do so*).[9] In sum, ACC proper functioning is important for what researchers call "motivated action"; as Damasio puts it, the ACC constitutes the "source for the energy of both external action (movement) and internal action (thought animation, reasoning)."[10]

Recent fMRI work by MacDonald et al. (2000) using a modified version of the Stroop task (which involves naming the ink color of colored words;
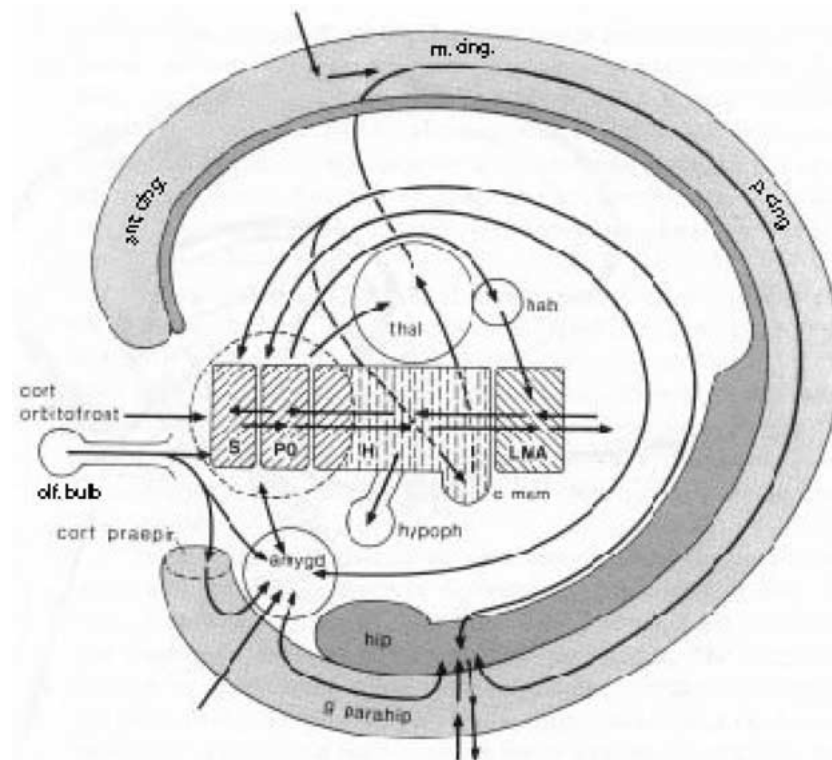
Figure 5. A schematic showing some of the pathways in the limbic structures. a. cing anterior cingulate gyrus; m. cing., middle cingulate gyrus; p. cing g. posterior cingulate gyrus; hip, hippocampus; g. parahipp, parahippocampal gyrus; thal, thalamus; S, septum; PO, preoptic region; H, hypothalamus; LMA, limbic midbrain area, c mam, mammilary body; hypoph, hypophysis (pituitary); hab, habenula; olf bulb, olfactory bulb, amygd, amygdala. Modified from Nieuwenhuys (1985).

famously, it takes subjects longer to name the color of the ink if the word *spells* a *differing color*) indicates that the ACC is involved in evaluative processes, "such as monitoring the occurrence of errors or the presence of response conflict, which occurs when two incompatible responses are both compelling."[11] They conclude that the ACC is particularly important for identifying just when it is that an organism needs to be more strongly engaged in controlling its behavior. Experiments by van Veen et al. (2001, 2002) support this conclusion, as does work by Bunge et al. (2001), which suggests the anterior cingulate is involved in detecting the organism's need for higher cognitive control. This role for the ACC is consistent with it being, in some obscure sense, a fountainhead for motivated action. When

the ACC is dysfunctional, subjects become catatonic, with severe impairment of reasoning and emotional expression, or they may become hyper-addicted, with similar impairments. Alien hand syndrome, in which the ipsilesional hand tends to behave independently of the agent's conscious control, is associated with ACC lesions that probably also involve the adjacent commissures. The alien hand may, for example, grab food from someone else's plate or grope a stranger, much to the embarrassment of the agent. Sometimes a person with alien hand syndrome is able to control the hand by hollering at it to stop.

Hippocampal structures
The hippocampal structures include the hippocampus, the parahippocampal cortex, the entorhinal cortex and the perirhinal cortex, all of which are organized in a very precise manner and show highly specific activity patterns in the awake period and during REM. The route from other parts of cortex to the hippocampus is through the entorhinal cortex. This circuitry is essential for learning and remembering specific events or episodes. These structures are not, however, the site of permanent memory storage, but are needed for the transfer of information to various parts of the cortex. It is still very unclear, however, what exactly the hippocampus does in processing events for storage, what exactly is transferred to cortex and how the transfer to cortex is achieved. According to the leading hypothesis (Hobson and Pace-Schott 2002), memory rehearsal occurs in the hippocampus during REM, and when the brain shifts to deep sleep stages, information is transferred from hippocampus to cortical structures. Some single neuron research on rats tends to support this hypothesis, as do the data exploring the regular cranking up and down of the neurotransmitter systems (noradrenalin, serotonin and acytelcholine) that regulate shifts from wakeness to REM to deep sleep.

Memory in electronic computers is absolutely the wrong model for memory in brains. In brains, information storage is essentially never the relevance-free, value-free, skill free, past-free business it is in electronic computers. In brains, not all details of an experience are stored and remembered, but generally only the highly processed and the "salient" or *me-relevant* parts. This seems to imply that in the weeks and months during which the hippocampus is "handling" such an input, a lot of sifting, sorting and reconstructing occurs. This business, whatever it really is, depends on many factors, including the brain's needs, its previously stored memories, its implicit memory, future plans, mood and so forth.

In moral judgment, it appears the hippocampus facilitates conscious recollection of schemas and memories that permit past events to figure in current decisions. For example, faced with a complex problem such as how best to

resolve a hostage situation, or how to persuade a child to acquire manners, the recollection of relevant past experiences along with their emotional valence is important, and this recollection is dependent on hippocampal structures, but also involves other limbic structures as well.

A caveat: many of the most important forms of moral reasoning that we exercise rely on background social skills, social perceptions and tacit use of a theory of mind. That is, much day-to-day moral reasoning does not consciously involve complex moral schemata or convoluted moral modeling. Consider, for example, airplane manners or doctor's waiting room manners: one does not have to deliberate about whether to stare at someone odd, whether to laugh at a pratfall of a shy child, whether to jump ahead of handi-capped passengers, or whether to pretend to be carrying a bomb. One does not have to deliberate about whether to protect a child from a snarling pit bull or from a careening car. Much of the time, we rely for fast judgments on well-worn skills and habits of character, informed by feelings.

Nonetheless, abstract moral reasoning *is* sometimes needed, particularly when our skills give conflicting judgments or no judgment, or judgments opposed by those we respect, and yet *something* must be done. And a decision *not* to act has consequences every bit as much as a decision *to* act. Should I give the order to put sleeping gas in the theatre under siege? Should I, as a member of the jury, agree to the death penalty of a child murderer? Should I tell a lie to protect an innocent person? Abstract moral reasoning probably depends on those brain structures that subserve morally-neutral forms of abstract thought as well as practical reasoning about how to get things done. Moral reasoning is one aspect of the ubiquitous real-world, biological, "dirty" computation that permits a brain to navigate a world and manage in a range of environments, new as well as familiar. This sort of computation gener-ally aims to get a good answer in good time, rather than an optimal answer in however-long-it-takes. Proving a theorem in mathematics might have no time pressure whatever; making a decision under threat from a human or other animal virtually always has severe time constraints. Dirty computation makes decisions in appropriate time under uncertainty, and it is guided by a range of feelings generated by current perceptions, past memories, imagined consequences, deep preferences, and passing fancies – all toned and molded by the brainstem's neuromodulator systems.

Moral state-spaces
A useful conceptual device for thinking about the moral nature of the activ-ities accomplished by these brain systems is to construe their activity as embodying a moral state space. We can usefully think of much of the activity of frontal cortex and the limbic/brainstem axis as consisting of a moving point
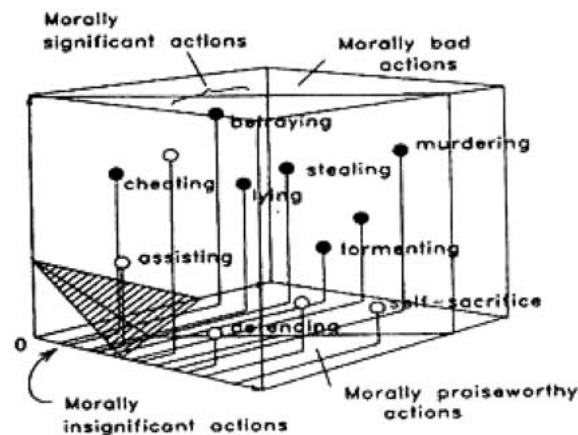
Figure 6. A hypothetical moral state space after dimensional reduction. Modified from Paul Churchland (1998).

in an *n*-dimensional space, where *n* could (in the most complex of cases) be determined by making the activity of every neuron involved in the subsystem an axis of the space. Using the mathematical technique of multi-dimensional reduction, it may be possible to boil these high-dimensional spaces into lower-dimensional ones wherein we can capture the principal components of the space using traditional conceptual terms. Alternately, our failure to capture such principal components areas might signal the need for articulation of a new vocabulary and concepts necessary to fully capture the dynamics and structure of the space. As an example it might be appropriate to think of the activity of these higher-order systems as taking place in this hypothetical moral state-space, and this can serve as a useful conceptual device.

Recap and coda: Ecological validity and the presence of unarticulated foundational assumptions

A problem with brain research on moral phenomena is that constraints of rigorous experimental design mean that tests are conducted in highly artificial situations that often do not resemble the actual circumstances in which real moral cognition occurs. "Ecological validity" is difficult achieve, especially in the case of moral reasoning. There are several reasons why this is so. Moral cognition is oft-times:

1. *Hot.* That is, affective and conative states are part and parcel of moral judgment (for example, the virtuous person is appropriately *angry* at unjust events). This is not an accidental feature, but an endemic feature.

Hot cognition is especially difficult to capture in artificial experimental settings.

2. *Social.* Moral judgments are virtually never made in a social vacuum. Rights, duties, obligations, norms of conduct, etc., all are made manifest and have their origin in a social atmosphere. Normally one lives among kin and community, where there are social expectations, social hierarchies, divisions of labor and conventions for dealing with those who violate social norms. This characterizes animal and well as human groups. Social environments are notoriously difficult to simulate when you are encased in a giant fMRI magnet, although Montague et al.'s "multi-scanner" methodology will go a long way towards correcting this deficiency.[12]

3. *Distributed.* Moral cognition is part of a broader network of understanding molded by multi-modal interactions with the world. Life is usually not limited to a single dimension of stimulation, but embedded in a rich context. Evolution is very conservative and new capacities are knit into preserved structures, not built from scratch. Social behavior appears to be rooted in the brainstem/limbic axis and prefrontal cortex, with input from sensory and multi-modal corteces. Notice that this embraces pretty much the whole brain. Consequently, the worry is that the optimal conditions for engaging the affective/cognitive equipment do not coincide with the reduced conditions needed to do experiments in an fMRI chamber (e.g., using bits of written text, small photographs, etc.).

4. *Organic.* Moral judgments are exquisitely context sensitive. Concerns about the fundamental attribution error aside, the fact remains that in *one* context I am likely to praise someone for telling a lie ("You look lovely in that spandex outfit"), and in another I am likely to *condemn* them ("Here at Enron our books are in *perfect* order"). Experimental setups will eventually have to find a way to take this context sensitivity into account, and that is difficult to do without compromising other aspects of the testing regime (sometimes, those very aspects that are necessary to get interpretable results at all).

5. *Genuine.* Many of the phenomena our moral theories and their psychologies hope to explain include things like *akrasia* ("weakness of the will"), moral motivation (when do people behave morally, and why?) and moral heroism, as when someone sacrifices his life for another. Addressing these phenomena requires exposing subjects to bona fide morally laden situations to engage the affective/cognitive functions. Experimental setups which assume too much regarding connections between dry "moral judgment" and wet "action in the world" may lead us

    to incorrect conclusions regarding the interface of emotion, reason, and action.

6. *Directed.* Anti-realism about morality aside, moral judgments are about *things*, broadly construed. In much the same way that a visual system seeks to construct a stable percept of the world, so does a moral cognitive system seek to construct a stable projection of the interface of cognition and action. In much the same way that the visual system becomes increasingly skilled in categorizing the causal structure of the world, so the social affective/cognitive system becomes increasingly skilled in navigating the social *cum* practical world. Moral judgments tell us what we ought to think *so that* we know what to do. Isolating the doing from the knowing via an artificial experimental regimen can remove the directedness of moral cognition. As for me and my house, when in the magnet, we will throw the switch to divert the trolley on to the track where only one child is playing. But what we'd do outside the magnet is likely anyone's guess.

Thus far, those experimental setups that are *most* ecologically sensitive support the hypothesis that moral cognition is a large-scale brain affair depending on the appropriate coordination of many areas. Research at a range of levels of organization from synapses to neurons to areas and systems points to the conclusion that the organism which best triangulates norms will be one that uses (1) multi-modal signals (2) conjoined with appropriately cued executive systems that (3) share rich connections with affective and conative brain structures (4) which draw upon conditioned memories (5) and insight into the minds of others so as to (6) think about and actually behave in a manner enabling it to function as best it can. These capacities will have neural correlates, but generally such correlates will be multi-faceted high-order functional relationships. Although *localization* of functions in cortex is still a common aim, increasingly the evidence shows that especially for complex functions, the aim is misguided. While it is true that there are regions that in the adult seem more or less specialized for particular functions, the idea that there are centers or modules or localized functions is largely a myth of pre-1980 neuroscience. If we cannot expect modules for depth-perception, motion perception or language, it is unwise to expect them for affective/cognitive representation.

Fair warning: even these seemingly commonplace observations about moral cognition smuggle in an agenda. When combined with these caveats, the research discussed so far indicates there probably is no such thing as a "pure reasoning" capacity (contra Kant), that purely cognitive moral judgments probably do not motivate intrinsically (contra Plato), that good moral cognition is shot-through with emotion (contra Kant), that moral judgment and social judgment are tightly intertwined (contra Kant), and that the ability

to accomplish utility calculations alone does not adequately capture the demands of good moral reasoning (contra Mill). We think the literature thus far best coheres with an appropriately updated and scientifically burnished virtue theory along the lines advocated by Aristotle in his classic treatise *Nicomachean Ethics*. These conclusions highlight the necessity of making the foundational assumptions that drive experimental setups an *explicitly articulated* portion of further work on the NMMC.

Moreover, they highlight the necessity of viewing moral cognition as being on a par with everyday forms of reasoning. For example, the question "Who should I dance with?" is not radically different from the question "Shall I tell a lie?" and (as one's limbic system knows) both can be relatively trivial or fraught with social peril. From the point of view of the brain and its me-relevant style of computation, there is no genuine distinction between "pure" practical reason and moral reasoning. Instead, at bottom it is only a matter of convention that we call certain judgments about what will be most functional for us "moral judgments" while others are purportedly *merely* judgments of prudence (although one difference is often that the former are highly affect-laden). One caveat, however: moral reasoning is reasoning about what one should think or do so as to manage life well (on our neo-Aristotelian paradigm), and it *is* true that some of our most significant reasoning in this domain deals with serious matters of social interaction and thus will reliably activate, for example, ToM modules and limbic "tags." There's room for amicable détente here.

Concluding thoughts

Until recently, moral philosophy has remained relatively detached from the natural sciences, including oddly enough, evolutionary biology. Neuroscientists and cognitive scientists, on the other hand, have recently become intensely interested in social/moral behavior as techniques have made its empirical exploration possible. What are needed are cross-disciplinary studies that rely on the best research of all the relevant disciplines. This means that moral philosophers will need to think the unthinkable, namely that the empirical results in the evolutionary and neural sciences have an unavoidable impact on moral philosophy. But it also means they can be useful in designing ecologically valid experiments. Likewise, it means scientists will need to develop approaches, both theoretical and experimental, to understand real-world, "dirty," *biological* computation and how it produces social and functional behavior. Far from being drudgery, these challenges are exciting – rarely have moral philosophers and cognitive scientists lived in such interesting, non-faddish, "justifiably hot" times.

Notes

[1] Bill would like to thank Dr. George Mastroianni and Dr. Kathleen Akins for their suggestions, and Dr. Jorge Moll for providing reprints of his team's research reports into the NMMC.
[2] See Sharon Begley's article 'Religion and the Brain', the cover story for *Newsweek*, May 7, 2001, or see her March 13, 2000 *Newsweek* item 'Learning Right from Wrong', for examples of popular press articles about 'moral neurobiology.' See also Oliver Goodenough's legal essay 'Mapping Cortical Areas Associated with Legal Reasoning and Moral Intuition' (2001). Goodenough is to be commended for bringing neurobiology to bear on legal issues. In that vein, see also Laura Reider's article 'Toward a New Test for the Insanity Defense: Incorporating the Discoveries of Neuroscience into Moral and Legal Theories' (1998).
[3] Moral philosopher G.E.M. Anscombe (1958) is famous for going so far as to issue a *moratorium* on normative moral theorizing until we have a better moral psychology in place. While this is a bit drastic (as it partially misconstrues the role that normative moral theory plays in moral psychology), we are nonetheless sympathetic and agree that moral psychology, broadly construed, will play a *critical* role in helping us triangulate on an appropriate norm-producing moral theory. For more on the impact of the cognitive and biological sciences in general on moral theory, see William Casebeer's forthcoming *Natural Ethical Facts: Evolution, Connectionism, and Moral Cognition* (The MIT Press).
[4] See his (1961) 'A Method for Studying Depth Perception in Infants under Six Months of Age', or his (1965) 'Visual Perception from Birth as Shown by Pattern Selectivity.'
[5] Here is an online resource for more information about Gage: http://www.deakin.edu.au/hbs/GAGEPAGE/.
[6] Gage even has two books about him: Malcolm Macmillan's *An Odd Kind of Fame: Stories of Phineas Gage* (2000), and John Fleischman's children's book (!) *Phineas Gage: A Gruesome but True Story About Brain Science* (2002).
[7] See Alison Gopnik's summary in her 'Theory of Mind' entry from the *MIT Encyclopedia of the Cognitive Sciences* (1999: 838–841).
[8] Paré et al. (2002) note that the amygdala projects to virtually all levels of the central nervous system; they postulate that amygdala oscillations during sleep may play a role in promoting synaptic plasticity in these portions of the nervous system, thus facilitating interactions between neocortical storage sites and temporal lobe structures involved in declarative memory. This purportedly explains why sleep deprivation harms memory recall.
[9] The technique used to capture fMRI data from two subjects or more simultaneously as they are interacting socially is called 'Hyperscanning'; this type of development promises to add another layer of ecological validity to the imaging regimen. We applaud it and hope to see more researchers using it, as we mention in section 4.0. See Montague et al. (*NeuroImage*, 2002).
[10] Damasio (1994: 71).
[11] MacDonald et al. (2000: 1836).
[12] See their (2002) article 'Hyperscanning: Simultaneous fMRI during Linked Social Interactions', *NeuroImage* **16**(4), 1159–1164, mentioned previously in footnote 9.

References

Adolphs, R.: 2002, 'Trust in the Brain', *Nature Neuroscience* **5**(3), March, 192–193.
Adolphs, R., Tranel, D. and Damasio, A.R.: 1998, 'The Human Amygdala in Social Hudgment', *Nature* **393**, June 4, 470–474.

Anderson, S.W., Bechara, A., Damasio, H., Tranel, D. and Damasio, A.R.: 1999, 'Impairment of Social and Moral Behavior Related to Early Damage in Human Prefrontal Cortex', *Nature Neuroscience* **2**(11), November, 1032–1037.

Anscombe, G.E.M.: 1958, 'Modern Moral Philosophy', *Philosophy* **33** (reprinted in (1997), R. Crisp and M. Slote (eds), *Virtue Ethics: Oxford Readings in Philosophy*, Oxford University Press, New York.

Baxter, M.G. and Murray, E.A.: 2002, 'The Amygdala and Reward', *Nature Reviews: Neuroscience* **3**, July, 563–573.

Bechara, A., Damasio, A.R., Damasio, H. and Anderson, S.W.: 1994, 'Insensitivity to Future Consequences Following Damage to Human Prefrontal Cortex', *Cognition* **50**, 7–15.

Bechara, A.: 2002, 'The Neurology of Social Cognition', *Brain* **125**, 1673–1675.

Berthoz, S., Armony, J.L., Blair, R.J.R. and Dolan, R.J.: 2002, 'An fMRI Study of Intentional and Unintentional (Embarrassing) Violations of Social Norms', *Brain* **125**, 1696–1708.

Brothers, L.: 1997, *Friday's Footprint: How Society Shapes the Human Mind*, Oxford University Press, New York.

Bunge, S.A., Ochsner, K.N., Desmond, J.E., Glover, G.H. and Gabireli, J.D.E.: 2001, 'Prefrontal Regions Involved in Keeping Information In and Out of Mind', *Brain* **124**(10), October, 2074–2086.

Call, J.: 2001, 'Chimpanzee Social Cognition', *Trends in Cognitive Science* **5**, 388–393.

Canessa, N., Gorinin, A., Perani, D., Danna, M., Scifo, P., Cappa, S.F., Piattelli Palmar-inin, M. and Fazio, F.: 2002, 'Social Cognition and Reasoning: Neuroimaging Evidence', *NeuroImage Human Brain Mapping 2002 Meeting*, abstract #1025.

Casebeer, W.D.: forthcoming, *Natural Ethical Facts: Evolution, Connectionism, and Moral Cognition*, The MIT Press, Cambridge, MA.

Caspi, A., McClay, J., Moffitt, T.E., Mill, J., Martin, J., Craig, I.W., Taylor, A. and Poulton, R.: 2002, 'Role of Genotype in the Cycle of Violence in Maltreated Children', *Science* **297**, 851–854.

Churchland, P.S.: 2002, *Brain-Wise: Studies in Neurophilosophy*, The MIT Press, Cambridge, MA.

Churchland, P.S.: 1995, 'Feeling Reasons', in A.R. Damasio, H. Damasio and Y. Christen (eds), *Decision-Making and the Brain*, Springer-Verlag, Berlin.

Churchland, P.S.: 1986, *Neurophilosophy: Towards a Unified Science of the Mind/Brain*, The MIT Press, Cambridge, MA.

Churchland, P.: 1998, 'Towards a Cognitive Neurobiology of the Moral Virtues', *Topoi* **17**, 83–96.

Coolidge, F.L. and Wynn, T.: 2001, 'Executive Functions of the Frontal Lobes and the Evolutionary Ascendancy of Homo Sapiens', *Cambridge Archaeological Journal* **11**, 2.

Cytowic, R.E.: 1996, *The Neurological Side of Neuropsychology*, The MIT Press, Cambridge, MA.

Damasio, A.R.: 1994, *Descartes' Error: Emotion, Reason, and the Human Brain*, G. P. Putnam and Sons, New York.

Damasio, A.R.: forthcoming, *Looking for Spinoza*, G. P. Putnam and Sons, New York.

Dolan, R.J.: 1999, 'On the Neurology of Morals', *Nature neuroscience* **2**(11), November 1999, 927–929.

The Economist: 2002, 'The Ethics of Brain Science: Open Your Mind'. May 23.

Fantz, R.L.: 1961, 'The Origins of Form Perception', *Scientific American* **204**, 66–72.

Fantz, R.L.: 1965, 'Visual Perception from Birth as Shown by Pattern Selectivity', *Annals of the New York Academy of Science* **118**, 793–814.

Fincham, J.M., Carter, C.S., van Veen, V., Stenger, V.A. and Anderson, J.R.: 2002, 'Neural Mechanisms of Planning: A Computational Analysis Using Event-related fMRI', *Proceedings of the National Academy of Sciences* **99**(5), March 5, 3346–3351.

Fleischman, J.: 2002, *Phineas Gage: A Gruesome but True Story About Brain Science*, Houghton Mifflin Co.

Fuster, J.M.: 1995, *Memory in the Cerebral Cortex: An Empirical Approach to Neural Networks in the Human and Nonhuman Primate*, The MIT Press, Cambridge, MA.

Gehring, W.J. and Willoughby, A.R.: 2002, 'The Medial Frontal Cortex and the Rapid Processing of Monetary Gains and Losses', *Science* **295**, March 22, 2279–2282.

Goodenough, O.R.: 2001, 'Mapping Cortical Areas Associated With Legal Reasoning and Moral Intuition', *Jurimetrics* **41**, Summer, J. 429.

Goldman-Rakic, P.S.: 1987, 'Circuitry of the Pre-Frontal Cortex and the Regulation of Behavior by Representational Memory', in F. Plum and V. Mountcastle (eds), *Higher Cortical Function: Handbook of Physiology*, American Physiological Society, pp. 373–417.

Gopnik, A.: 1999, 'Theory of Mind', in R.A. Wilson and F.C. Keil (eds), *The MIT Encyclopedia of the Cognitive Sciences*, The MIT Press, Cambridge, MA, pp. 838–841.

Greene, J.D., Sommerville, R.B., Nystrom, L.E., Darley, J.M. and Cohen, J.D.: 2001, 'An fMRI Investigation of Emotional Engagement in Moral Judgment', *Science* **293**, September 14.

Hamann, S.B., Ely, T.D., Hoffman, J.M. and Kilts, C.D.: 2002, 'Ecstasy and Agony: Activation of the Human Amygdala in Positive and Negative Emotion', *Psychological Science* **13**(2), March.

Hobson, A. and Pace-Schott, E.F.: 2002, 'The Cognitive Neuroscience of Sleep: Neuronal Systems, Consciousness, and Learning', *Nature Reviews: Neuroscience* **3**, 679–693.

Houde, O., Zago, L., Crivello, F., Moutier, S., Pineau, A., Mazoyer, B. and Tzourio-Mazoyer, N.: 2001, 'Access to Deductive Logic Depends on a Right Ventromedial Prefrontal Area Devoted to Emotion and Feeling: Evidence from a Training Paradigm', *NeuroImage* **14**, 1486–1492.

Johnson, M.H. and Morton, J.: 1991, *Biology and Cognitive Development: The Case of Face Recognition*, Blackwell, New York.

Johnson, S.C., Baxter, L.C., Wilder, L.S., Pipe, J.G., Heiserman, J.E. and Prigatano, G.P.: 2002, 'Neural Correlates of Self-reflection', *Brain* **125**(8), 1808–1814.

Kahn, D., Pace-Schott, E. and Hobson, J.A.: 2002, 'Emotion and Cognition: Feeling and Character Identification in Dreaming', *Consciousness and Cognition* **11**, 34–50.

Laakso, M.P., Vaurio, O., Koivisto, E., Savolainen, L., Eronen, M., Aronen, H.J., Hakola, P., Repo, E., Soininen, H. and Tiihonen, J.: 2001, 'Psychopathy and the Posterior Hippocampus', *Behavioural Brain Research* **118**, 187–193.

MacDonald, A.W. III, Cohen, J.D., Stenger, V.A. and Carter, C.S.: 2000, 'Dissociating the Role of the Dorsolateral Prefrontal and Anterior Cingulate Cortex in Cognitive Control', *Science* **288**, June 9, 1835–1838.

Macmillan, M.: 2000, *An Odd Kind of Fame; Stories of Phineas Gage*, The MIT Press, Cambridge, MA.

Manes, F., Sahakian, B., Clark, L., Rogers, R., Antoun, N., Aitken, M. and Robbins, T.: 2002, 'Decision-making Processes Following Damage to the Prefrontal Cortex', *Brain* **125**(3), March, 624–639.

McCabe, K., Houser, D., Ryan, L., Smith, V. and Trouard, T.: 2001, 'A Functional Imaging Study of Cooperation in Two-person Reciprocal Exchange', *Proceedings of the National Academy of Sciences* **98**(20), September 25, 11832–11835.

Miller, G.: 2002, 'Gene's Effect Seen in Brain's Fear Response', *Science* **297**, July 19.

Miller, G.: 2002, 'The Good, the Bad, and the Anterior Cingulate', *Science* **295**, March 22, 2193–2194.

Moll, J., Slinger, P.J. and de Oliveira-Souza, R.: 2001, 'Front Polar and Anterior Temporal Cortex Activation in a Moral Judgment Task', *Arm. Neuropsiquiatry* **59**(3-B), 657–664.

Moll, J., de Oliveira-Souza, R., Eslinger, P.J., Bramati, I.E., Mourao-Miranda, J., Andreiuolo, P.A. and Pessoa, L.: 2002, 'The Neural Correlates of Moral Sensitivity: A Functional Magnetic Resonance Imaging Investigation of Basic and Moral Emotions', *The Journal of Neuroscience* **22**(7), April 1, 2730–2736.

Moll, J., de Oliveira-Souza, R., Bramati, I.E. and Grafman, J.: 2002, 'Functional Networks in Emotional Moral and Non-Moral Social Judgments', unpublished manuscript.

Montague, P.R., Berns, G.S., Cohen, J.D., McClure, S.M., Pagnoni, G., Dhamala, M., Wiest, M.C., Karpov, I., King, R.D., Apple, N. and Fisher, R.E.: 2002, 'Hyperscanning: Simultaneous fMRI during Linked Social Interactions', *NeuroImage* **16**(4), August 1, 1159–1164.

Nichols, S.: 2002, 'Norms with Feeling: Towards a Psychological Account of Moral Judgment', *Cognition* **84**, 221–236.

Oliveira-Souza, R. and Moll, J.: 2000, 'The Moral Brain: A Functional MRI Study of Moral Judgment', *Neurology 2000* **54**(7), A104.

Pare, D., Collins, D.R. and Pelletier, J.G.: 2002, 'Amygdala Oscillations and the Consolidation of Emotional Memories', *Trends in Cognitive Sciences* **6**(7), July, 306–314.

Paulus, M.P., Hozack, N., Zauscher, B., McDowell, J.E., Frank, L., Brown, G.S. and Braff, D.L.: 2001, 'Prefrontal, Parietal, and Temporal Cortex Networks Underlie Decision-Making in the Presence of Uncertainty', *NeuroImage* **13**, 91–100.

Peoples, L.L.: 2002, 'Will, Anterior Cingulate Cortex, and Addiction', *Science* **296**(31), May, 1623–1624.

Preston, S.D. and de Waal, F.B.M.: forthcoming, 'Empathy: Its Ultimate and Proximate Bases', *Behavioral and Brain Sciences* (available as a preprint at http://www.bbsonline.org/Preprints/Preston/Referees/)

Reider, L.: 1998, 'Toward a New Test for the Insanity Defense: Incorporating the Discoveries of Neuroscience into Moral and Legal Theories', *UCLA Law Review* **46**, 289.

Rilling, J.K., Gutman, D.A., Zeh, T.R., Pagnoni, G., Berns, G.D. and Kilts, C.D.: 2002, 'A Neural Basis for Social Cooperation', *Neuron* **35**, 395–405.

Rizzolatti, G., Fogassi, L. and Gallese V., 2001, 'Neurophysiological Mechanisms Underlying the Understanding and Imitation of Action', *Nature Reviews: Neuroscience* **2**, 661–670.

Shidara, M. and Richmond, B.J.: 2002, 'Anterior Cingulate: Single Neuronal Signals Related to Degree of Reward Expectancy', *Science* **296**, 31 May, 1709–1711.

Stich, S.: 1993, 'Moral Philosophy and Mental Representation', in M. Hechter, L. Nadel and R.E. Michod (eds), *The Origin of Values*, Aldine de Gruyter, New York, pp. 215–228.

Stone, V.E., Cosmides, L., Tooby, J., Kroll, N. and Knight, R.T.: 2002, 'Selective Impairment of Reasoning about Social Exchange in a Patient with Bilateral Limbic System Damage', *Proceedings of the National Academy of Sciences* **99**(17), August 20, 11531–11536.

Van Veen, V. and Carter, C.S.: 2002, 'The Timing of Action-Monitoring Processes in the Anterior Cingulate Cortex', *Journal of Cognitive Neuroscience*, **14**, 593–602.

Van Veen, V., Cohen, J.D., Botvinick, M.M., Stenger, V.A. and Carter, C.S.: 2001, 'Anterior Cingulate Cortex, Conflict Monitoring, and Levels of Processing', *NeuroImage* **14**(6), December, 1302–1308.

Waldbauer, J.R. and Gazzaniga, M.S.: 2001, 'The Divergence of Neuroscience and Law', *Jurimetric J.* **41**, Spring, 357.

194

Wallis, J.D., Anderson, K.C. and Miller, E.K.: 2001, 'Single Neurons in Prefrontal Cortex Encode Abstract Rules', *Nature* **411**, June 21, 953–956.

Winston, J.S., Strange, B.A., O'Doherty, J. and Dolan, R.J.: 2002, 'Automatic and Intentional Brain Responses during Evaluation of Trustworthiness of Faces', *Nature Neuroscience* **5**(3), March, 277–283.

Yamasaki, H., LaBar, K.S. and McCarthy, G.: 'Dissociable Prefrontal Brain Systems for Attention and Emotion', *Proceedings of the National Academy of Sciences* **99**(17), August 20, 11447–11451.