OPINION

# The neural basis of human moral cognition

*Jorge Moll, Roland Zahn, Ricardo de Oliveira-Souza, Frank Krueger and Jordan Grafman*

Abstract | Moral cognitive neuroscience is an emerging field of research that focuses on the neural basis of uniquely human forms of social cognition and behaviour. Recent functional imaging and clinical evidence indicates that a remarkably consistent network of brain regions is involved in moral cognition. These findings are fostering new interpretations of social behavioural impairments in patients with brain dysfunction, and require new approaches to enable us to understand the complex links between individuals and society. Here, we propose a cognitive neuroscience view of how cultural and context-dependent knowledge, semantic social knowledge and motivational states can be integrated to explain complex aspects of human moral cognition.

At a time of increasing awareness of the different value systems in multicultural societies and across nations, a deeper understanding of the cognitive and brain mechanisms that guide human behaviour is of general interest. Recent social cognitive neuroscience reviews have emphasized perceptual and emotional abilities that are shared by humans and other animals[1–3]. However, social neuroscience has largely avoided dealing directly with the complex aspects of human moral cognition, including MORAL EMOTIONS and MORAL VALUES. Here, we review current theoretical accounts of social cognition and put forth a framework designed to overcome the main limitations of earlier accounts. We argue that moral phenomena emerge from the integration of contextual social knowledge, represented as event knowledge in the prefrontal cortex (PFC); social semantic knowledge, stored in the anterior and posterior temporal cortex; and motivational and basic emotional states, which depend on cortical–limbic circuits. Our framework offers new interpretations for social behaviour patterns in healthy individuals and in patients with brain dysfunction, and makes testable predictions for neuropsychological dissociations in moral cognition.

## Defining morality

'Moral' (derived from the Latin *moralis*) and 'ethical' (from the Greek *êthikos*) originally referred to the consensus of manners and customs within a social group, or to an inclination to behave in some ways but not in others[4]. Through the centuries, philosophical theories have adopted a deductive logico–verbal approach to morality that aims to identify universal principles that should guide human conduct. By contrast, a scientific approach to morality is emerging from the documentation of changes in moral behaviour in patients with brain dysfunction[5], which provides inferences that concern the major dimensions of moral cognition. Moral cognitive neuroscience, therefore, aims to elucidate the cognitive and neural mechanisms that underlie moral behaviour. Here, morality is considered as the sets of customs and values that are embraced by a cultural group to guide social conduct, a view that does not assume the existence of absolute moral values. The implications of cognitive neuroscience for moral philosophy have been reviewed in detail elsewhere[6–8] and are not addressed here.

## The challenge

Morality is a product of evolutionary pressures that have shaped social cognitive and motivational mechanisms, which had already developed in human ancestors, into uniquely human forms of experience and behaviour[9]. Non-human primates have a vast repertoire of social behaviours that can be interpreted as genuine forerunners of human morality, such as caring for their peers and constantly striving for dominance[10]. As in humans, a sense of justice permeates their behaviour[11]. The evolution of the human PFC is intimately related to the emergence of human morality[12–15]. This has allowed motivational mechanisms to be integrated with an exceptional power to predict outcomes, and has characterized humans through their recent evolutionary steps in the cultural explosion of the Upper Paleolithic period[16].

The challenge for moral cognitive neuroscience is that it requires extensive cross-field integration of neuroscience, psychology, evolutionary biology and anthropology, among other areas. In setting the goals of scientific exploration in this field, some central issues should be considered. How does the human moral mind emerge from the interaction of biological and cultural factors? How can the context-dependent nature of moral cognition be explained by neuroscience? How does moral cognition relate to emotion and motivation, and what are their neural substrates? Although moral cognitive neuroscience is still in its infancy, the available evidence already points to some promising solutions.

## The neural basis of moral cognition

***Moral behaviour impairment.*** Persistent antisocial behaviours have long been described[17], yet their history in medicine is relatively recent. Impairment in 'moral sense', or 'moral insanity', was first formally described as a "perversion of natural feelings, affections,

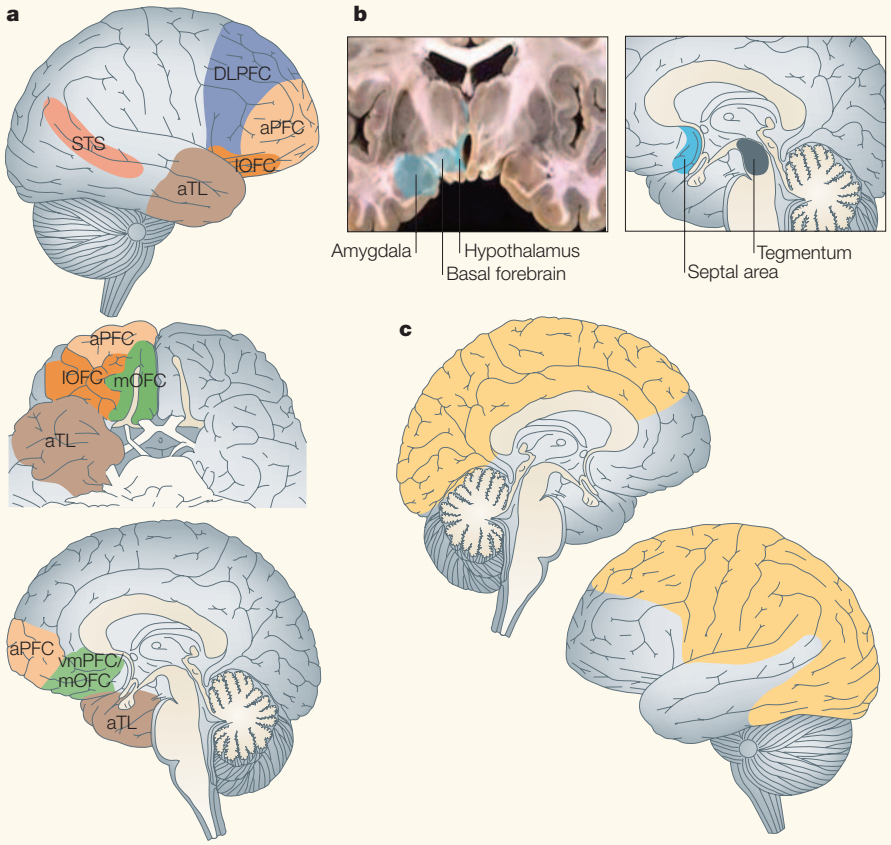



Figure 1 | **Brain regions implicated in moral cognition and behaviour in functional imaging and patient studies. a** | Cortical regions[13,99,107] include the anterior prefrontal cortex (aPFC), the medial and lateral orbitofrontal cortex (mOFC and lOFC), the dorsolateral PFC (DLPFC; mostly the right hemisphere) and additional ventromedial sectors of the PFC (vmPFC), the anterior temporal lobes (aTL) and the superior temporal sulcus (STS) region. **b** | Subcortical structures[13,36,48] include the amygdala, ventromedial hypothalamus, septal area and nuclei, basal forebrain (especially the ventral striatum/pallidum and extended amygdala), the walls of the third ventricle and rostral brainstem tegmentum. **c** | Brain regions that have not been consistently associated with moral cognition and behaviour in patient studies include the parietal and occipital lobes, large areas of the frontal and temporal lobes, the brain stem, basal ganglia and additional subcortical structures. Panel **b** modified, with permission, from REF. 147 © (2005) University of Iowa's Virtual Hospital. Anatomical image adapted, with permission, from REF. 148 © (1996) Appleton & Lange.

inclinations, temper, habits, moral dispositions, and natural impulses"[18,19]. Systematic evidence that specific brain regions might be crucial to moral behaviour was provided by early accounts of frontal lobe damage[20,21] and neurosurgical reports of war wounds (see, for example, REF. 22) (FIG. 1).

More recently, researchers have started to explicitly frame these observations within the sphere of moral cognition, strengthening the links between neuroscience, developmental neuropsychology and moral psychology. Eslinger and Damasio[23] described moral behavioural deficits in a patient with damage to the ventromedial PFC acquired in adulthood, who was remarkably unimpaired in specific MORAL REASONING tasks. It was later shown that ventromedial PFC lesions acquired at an early age led to impairments in both moral reasoning and behaviour, indicating that moral development can be arrested by early PFC damage[24,25]. These impairments in moral conduct resemble those observed in developmental PSYCHOPATHY[26,27] (BOX 1). Less frequently[28–31], lesions of the dorsolateral PFC (DLPFC; typically of the right hemisphere) also lead to changes in moral behaviour.

In addition to the PFC, other brain regions are crucial for moral cognition. Structural changes in the anterior temporal lobes — either acquired or developmental — can also impair moral behaviours[28,32]. Dysfunction of neural circuits that involve the superior temporal sulcus (STS) region — a key area for social perception[33] — is associated with the difficulty experienced by individuals with autism in attributing intentionality, which leads to reduced experience of pride and embarrassment[1,34]. Lesions to limbic and paralimbic structures can impair basic motivational mechanisms, such as sexual drive, social attachment and aggressiveness, leading to extreme moral violations — for example, unprovoked physical assaults and paedophilia[35,36]. Structural and functional imaging studies in psychopathic individuals have pointed to abnormalities in almost all these regions[37–40].

***Moral emotion and judgement.*** Recent studies have directly addressed the neural correlates of moral emotions and judgements. Patients with focal damage to the ventromedial PFC show deficient engagement of pride, embarrassment and regret[41,42]. Functional imaging studies in healthy individuals have involved simple MORAL JUDGEMENTS[43–45], moral dilemmas[46,47] and moral emotions[48,49–52], using different tasks and stimulus presentation schemes. Overall, there is remarkable agreement between functional imaging and clinico-anatomical evidence about the brain areas involved in moral cognition. Activated regions include the anterior PFC (encompassing the frontopolar cortex, Brodmann's area (BA) 9/10), orbitofrontal cortex (OFC; especially its medial sector, BA 10/11/25), posterior STS (BA 21/39), anterior temporal lobes (BA 20/21/38), insula, precuneus (BA 7/31), anterior cingulate cortex (ACC, BA 24/32) and limbic regions. Notably, the wide range of modalities, stimuli and task requirements appear to have little effect on brain activation patterns (FIG. 2).

Besides the consistent patterns of brain activation found across studies, there were also some differential findings. We found activation of the anterior PFC when a moral judgement condition was compared with non-emotional factual judgements[43], but not when moral judgements were compared with a social–emotional condition, during which a more ventral region was activated[44]. Greene and colleagues used a moral judgement task that involved classic moral dilemmas (for example, should you kill an innocent person in order to save five other people?) and found similar activation of the anterior PFC[46,47]. Decision difficulty was correlated with increased activity in the ACC. Heekeren and colleagues showed that the presence of bodily harm in moral violation scenarios leads to decreased reaction times and decreased activation of the anterior temporal lobe[53]. Evidence is emerging that partially dissociable PFC–temporal–limbic networks represent distinct moral emotions, including guilt, anger and embarrassment[13,49–52].

Box 1 | **Psychopathy and the neural organization of morality**

**The concepts of antisocial personality disorder ('sociopathy') and psychopathy (a severe form of sociopathy) originated from the need to diagnose individuals who show a pattern of behaviours that goes against the common good and repeatedly involves harm to others. Although social norms vary among cultures and even among intracultural niches, sociopathy and psychopathy cannot be reduced to 'cultural artefacts'[141] for the simple fact that their core manifestations are stable and easily recognizable, both historically and cross-culturally. The neurobiological validity of sociopathy and psychopathy is supported by increasing scientific evidence that the brains of affected individuals differ from those of socially adjusted people: imaging studies in psychopaths have revealed reduction of grey matter in the prefrontal cortex and abnormal brain activation in limbic regions, as well as in the prefrontal and temporal lobes[38,39].**

## Current accounts

Some current cognitive neuroscience frameworks have direct implications for our understanding of the neural basis of moral cognition. The main characteristics and limitations of these accounts are briefly reviewed and discussed below, with an emphasis on their relevance to moral cognition (see also TABLE 1).

*Conflict processing in moral judgement.* On the basis of functional imaging studies[46,47], Greene and colleagues have focused on the role of cognitive control in moral judgement. Their hypothesis was derived in part from Miller and Cohen's theoretical account of PFC function[54], which assumes that the PFC is specifically involved in 'controlled processing', such as in rapidly changing, ill-structured situations — characteristics that are also held by other models[14]. This proposal is supported by evidence for DLPFC and ACC activation in response to increases in attentional and conflict detection demands. Greene's hierarchical processing view assumes that cognitive control processes, afforded by the lateral PFC and ACC ('cognitive areas'), override emotional responses (which are attributed to the medial PFC, posterior cingulate cortex and STS) to produce UTILITARIAN responses to moral dilemmas — for example, smothering a crying baby to save more lives. By contrast, emotional areas would favour 'personal' moral judgements — for example, thinking that it is inappropriate to smother the baby. The theory posits mutually competitive roles of cognition and emotion in moral judgement.

Greene's functional imaging findings are in line with the cognitive control view and demonstrate reliable task-related effects in different types of moral judgement. However, the concepts of 'personal' and 'impersonal' violations, and of 'utilitarian' and 'non-utilitarian' choices need to be broken down into clear cognitive components. Furthermore, this account does not address the possibility that culturally shaped moral values and beliefs might lead to disparate 'utilitarian' conclusions. Finally, it is unclear how impairment of cognitive control and conflict monitoring would affect moral cognition[55].

*Somatic marker hypothesis.* Damasio and colleagues observed that patients with ventromedial PFC damage can detect the implications of a social situation, but cannot make appropriate decisions in real life. They suggested that such patients would be unable to mark those implications with a signal that automatically distinguishes advantageous from pernicious actions[56]. The somatic marker model explains why patients with ventromedial PFC damage can still reason about social problems, provided the premises are cast verbally, but fail in natural settings. The IOWA GAMBLING TASK, which was preceded by similar gambling tasks[57], was put forward

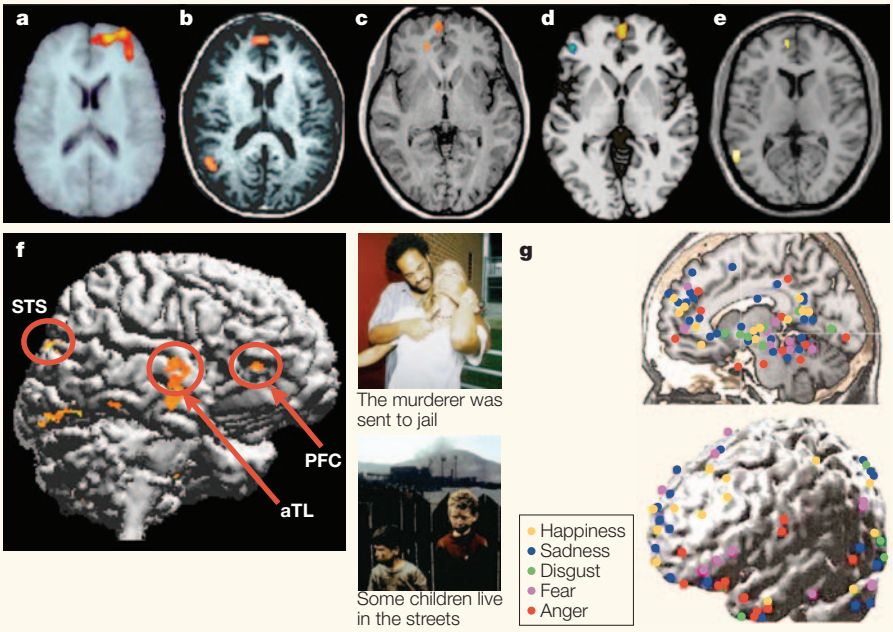



Figure 2 | **Functional imaging studies of moral cognition.** Functional imaging studies of moral cognition have revealed consistent involvement of the anterior prefrontal cortex (aPFC) and superior temporal sulcus (STS) region, as well as the anterior temporal lobes (aTL) and limbic structures. Panels **a**–**e** depict a transverse slice showing the activation of the aPFC (frontopolar cortex, Brodmann's areas 9/10) across different studies[43,46,48,45,50]. Panel **f** shows spatially overlapping activations in the PFC, STS region and aTL, derived from a conjunction analysis of two different studies: active moral judgements of written stimuli[44] and passive viewing of pictures with moral content[48]. Samples of the pictorial[149] and written stimuli used in these studies are shown. The remarkable overlap of brain regions involved in moral cognition, regardless of a wide variation in task requirements and stimulus modalities, contrasts with the large variability observed in brain imaging studies of 'less complex' basic emotions depicted in panel **g** (REF. 150). A strong possible explanation is the effect of familiarity and situational context, which have not been controlled in functional imaging studies of basic emotions or moral cognition. The higher reproducibility of the activation patterns in studies of moral cognition might, therefore, have resulted from a smaller contextual variability related to the use of more well-defined social situations for moral judgements and moral emotions. By contrast, experimental designs of studies of basic emotion have put more effort into equating the sensory properties of stimuli (such as luminance, visual complexity and frequency) at the cost of more variability in social contexts (such as fear associated with a picture of a spider or with a crime scene). Panel **a** reproduced, with permission, from REF. 43 © (2001) Associacao Arquivos de Neuro-Psiquiatria. Panel **b** reproduced, with permission, from REF. 46 (2001) American Association for the Advancement of Science. Panel **c** reproduced, with permission, from REF. 48 © (2002) Society for Neuroscience. Panel **d** reproduced, with permission, from REF. 45 © (2003) Lippincott Williams and Wilkins. Panel **e** reproduced, with permission, from REF. 50 © (2004) Elsevier Science. Panel **f** (right-hand images) from REF. 149. Panel **g** reproduced, with permission, from REF. 150 © (2002) Elsevier Science.

Table 1 | **Characteristics and limitations of frameworks relevant to moral cognitive neuroscience**

| Theoretical accounts | Situational and cultural context effects | Relationship between cognition and emotion | Predictions from brain lesions | Implications for moral cognitive phenomena | Key references |
|---|---|---|---|---|---|
| Conflict control in moral judgement | No predictions | Hierarchical; emotions inhibited by cognitive processes | No | Task difficulty and cognitive control in moral judgement | 47,54 |
| Somatic marker hypothesis | Partially addressed | Integrated; autonomic signals guide cognition | Yes | Implicit assessment of decision outcomes | 56,58,59 |
| Social response reversal | Partially addressed | Hierarchical; emotional signals help suppress aggression | Yes | Suppression of aggressive behaviour | 62,69,70 |
| Impairment of ToM mechanisms in sociopathy | No predictions | No predictions | Yes | Impairments in attribution of mental states in sociopathy | 76–78 |
| Structured-event-complex framework | Predicted | Hierarchical; PFC representations control emotional responses | Yes | Contextual effects and social knowledge | 14,15,84 |
| Moral sensitivity hypothesis | Partially addressed | Integrated; social perception and moral emotions are bound together | Yes | Moral emotions and implicit moral appraisals | 13,43,44,48 |
| Event–feature–emotion complex framework | Predicted | Integrated; social features, motivation, emotion and contextual knowledge are bound together | Yes | Binding of context-dependent social knowledge, features and emotion | 13,48,82,87 |

PFC, prefrontal cortex; ToM, theory of mind.

as an experimental surrogate for decision-making in real life. Bechara and colleagues[58] showed that normal individuals develop anticipatory galvanic skin responses whenever they contemplate a risky choice, and begin to choose advantageously before they are consciously aware of the best strategy. Patients with ventromedial PFC damage do not develop anticipatory autonomic responses and behave as if they are insensitive to future consequences, positive or negative, being primarily guided by immediate prospects that ultimately lead to a net financial loss.

The somatic marker hypothesis has been influential and is considered to be a possible mechanism that could underlie behavioural dysfunction in patients with PFC lesions. This framework is compatible with contextual effects (although these are not explicitly addressed), integrates cognition and emotion, makes testable predictions, and has been supported by neurophysiological and clinical data[58–60]. However, it does not explicitly address the role of different PFC subregions in moral cognition. The relationships between somatic markers and other cortical and limbic regions that have previously been linked to moral cognition[13] are also obscure. Recent evidence from both patients with PFC lesions and healthy individuals has challenged the role of somatic markers in guiding decision making and social behaviours[61–63].

***Social response reversal.*** The social RESPONSE-REVERSAL model, which was proposed by Blair and Cipolotti to explain social behavioural impairments in patients with OFC damage, was influenced by Rolls and coworkers' response-reversal paradigm. In their pioneering work, Rolls and colleagues showed that patients with OFC damage were impaired in EXTINCTION and response-reversal tasks[64]. These impairments were correlated with measures of socially inappropriate behaviours, which led to the hypothesis that the sociopathy of these patients results from a difficulty in modifying behavioural responses, especially when these are followed by negative outcomes. The response-reversal model has received extensive support from electrophysiological studies in animals[65] as well as from human lesion and neuroimaging data[66–68].

Blair and Cipolotti compared their findings from a patient with OFC damage (J.S.) with those from a patient with DLPFC damage and five prison inmates with psychopathy[62]. J.S. showed a drastic change in personality after OFC damage, becoming aggressive and callous towards other people. He was impaired in recognizing facial expressions of anger and disgust, but was unimpaired in response-reversal tasks. This led the authors to argue for a social response-reversal mechanism — an inhibitory system reliant on the proper functioning of the OFC that is normally activated by perception or expectation of others' anger[69]. Blair suggested that a different inhibitory mechanism — the 'violence inhibition mechanism' (VIM) — would be deficient in developmental psychopathy, leading to instrumental aggression[70]. The VIM underscores the role of the amygdala in aversive conditioning, and is believed to have a key role in moral socialization.

These accounts can be used to make specific predictions about the role of response reversals and aversive conditioning in patients with OFC and amygdala lesions. However, they cannot be easily extended to explain other types of impairment in moral behaviour that arise from damage to other brain regions, such as the temporal lobes and anterior PFC. In addition, these models were not designed to explain how social knowledge, on which reinforcement contingencies operate, is represented in the brain. Finally, although bilateral amygdala lesions lead to impaired perceptual judgement of facial emotions[71], evidence for severe impairments in moral behaviour following isolated amygdala lesions acquired either in adulthood or early childhood is still lacking.

***Sociopathy as a failure of 'theory of mind'.*** Disruptive antisocial behaviour is a hallmark of early frontotemporal dementia[72]. These profound changes in personality have been predominantly ascribed to degeneration of

the right PFC[73] or the temporal poles[74,75]. Lough *et al.*[76] used a battery of neuropsychological and social cognition tests to assess J.M., a 47-year-old man who presented with a decline in work performance and a gross deterioration in social behaviour. Imaging studies revealed bilateral atrophy of the OFC and anterior temporal lobes, including the amygdala. J.M. had a normal IQ and fared well on standard executive tests, but was otherwise severely impaired on THEORY OF MIND (ToM) tasks that require a degree of abstraction, with specific deficits on first- and SECOND-ORDER FALSE BELIEF TASKS, and on detection of *faux pas*. The authors proposed that the dissociation between the impairment in ToM mechanisms and normal executive performance underlies the personality changes observed in some cases of frontotemporal dementia. This account is therefore compatible with abnormal moral cognition — such as difficulties in the attribution and experience of pride and embarrassment — observed in autism and Asperger's syndrome, which are typically associated with ToM impairments[34,77]. However, ToM abilities only account for some aspects of moral cognition, but not, for example, the role of social knowledge, contextual information and basic motivations. Noticeably, ToM is relatively intact in psychopathy, in line with its role in the deviousness of these individuals[78].

***Structured-event-complex framework.*** The structured-event-complex (SEC) framework[15] supports claims that executive functions performed by the PFC are based on stored event sequence knowledge. SEC representations are long-term memories of event sequences that guide the perception and execution of goal-oriented activities, such as going to a concert or giving a dinner party[79]. A SEC representation includes situational knowledge abstracted across events (concert) and the temporal organization of events (making a reservation, dressing up, and so on). Activated SECs sequentially bind representations of objects, actions and spatial maps stored in posterior brain regions. The SEC framework predicts that different subdivisions of the PFC store different types of content or domains of event knowledge[14,80,81]. Clinical and neuroimaging evidence supports this prediction, showing that different PFC regions are involved in representing social and emotional SECs (ventromedial PFC)[82,83], novel or multitasking event sequences (anterior PFC)[84,85] or overlearned sequences (more posterior PFC regions)[86,87]. The importance of the PFC for goal-oriented activities is also corroborated by recent functional imaging studies of future reward prediction[88].

Although this framework has clear implications for moral cognition, these rely on the hypothesis that the PFC stores the situational and temporal context of social knowledge. The SEC framework does not predict how PFC regions interact with limbic areas and other cortical regions to give rise to a range of moral cognitive phenomena, such as moral values and moral emotions.

> "Ecological validity is especially relevant for moral cognition studies, because moral cognition depends strongly on situational and cultural context."

***Moral sensitivity hypothesis.*** A final account is that of the moral sensitivity hypothesis[13,48]. Using a task that engaged participants as observers, we showed that the viewing of pictures that depicted moral violations specifically activated the anterior PFC, medial OFC, STS region, brainstem and limbic structures. Scenes associated with BASIC EMOTIONS (disgust and fear) activated similar brainstem and limbic regions (including the amygdala), but not the medial OFC and STS. These findings are consistent with the hypothesis that a network involving the anterior PFC, OFC, STS and limbic regions represents social–emotional events linked to 'moral sensitivity' — an automatic tagging of ordinary social events with moral values. This hypothesis was supported by the finding that the medial OFC, anterior PFC, STS and precuneus show increased coupling in a functional connectivity analysis[48], and by the observation that a similar set of regions is involved in moral reasoning and social perception. Although we proposed that the OFC is more involved in automatic social–emotional associations and that the anterior PFC has a role in predicting future social outcomes, the role of the PFC in context-dependent social situations was not addressed. In addition, the moral sensitivity hypothesis makes no predictions about specific impairments in moral cognition following selective damage to the anterior temporal lobes, the STS region and PFC subregions.

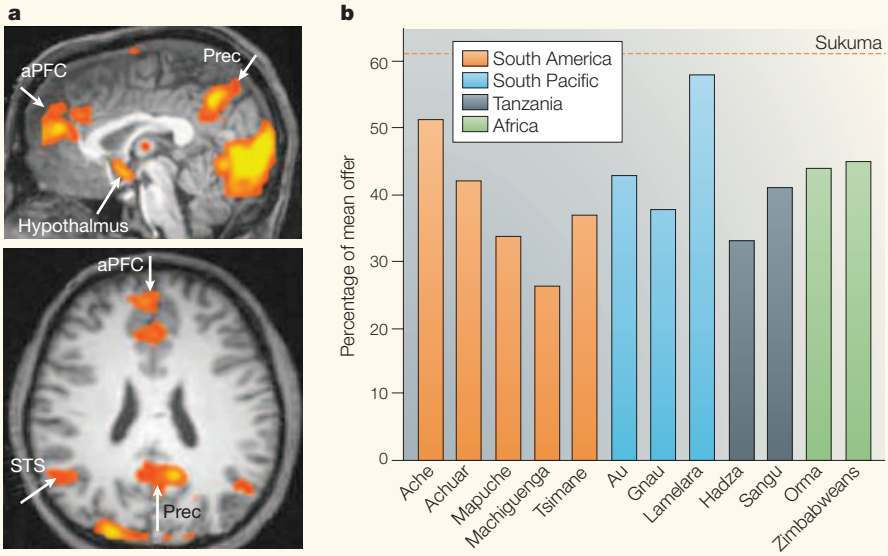
## Limitations of current frameworks

Some of the above frameworks point to clear-cut singular mechanisms. These mechanisms have the potential advantage of allowing more specific predictions to be made about the workings of particular brain regions, but they fall short of explaining key aspects of moral cognition. Some general limitations that apply to all of these frameworks are discussed below.

***Ecological validity of experimental designs.*** Ecological validity is especially relevant for moral cognition studies, because moral cognition depends strongly on situational and cultural context[6]. The experimental constraints that are imposed by behavioural and functional imaging studies might have an important impact on performance on moral cognition tasks. Some people might feel uncomfortable disclosing their opinions about sensitive issues, providing socially desirable answers instead. On the other hand, different people might provide similar opinions, but rely on entirely different moral values. The fact that moral cognition operates to a large extent swiftly and implicitly in regular social life[13] makes the ecological validity issue even more crucial. The making of moral judgements on extreme and unfamiliar situations, such as those posed by classic moral dilemmas[89], offers interesting ways to probe philosophical points of view, but can hardly be taken as a proxy for everyday moral reasoning. In addition, personal beliefs and familiarity with the scenarios strongly affect behaviour and brain activation results[90–94].

***Brain processes and representations.*** Another important limitation of current accounts is the lack of specific predictions about the effects of PFC lesions on moral behaviour. PFC function has been described using two general views: the 'processing' approach, which holds that the cognitive function of the PFC can be described in terms of performance without specifying a representation, and the 'representational' approach, which seeks to establish what type of information is stored in the PFC[14]. The processing view tends to regard the PFC as a content-free repository of processing modules, such as conflict monitoring, selection and inhibitory control[95], and predicts task-dependent rather than content-dependent dissociations resulting from brain lesions. Therefore, moral behaviour impairments following PFC damage would result from a release of limbic areas from PFC 'executive control'[96]. However, there is no convincing

Box 2 | **Culture, moral values and neuroeconomics**




**Humans often show altruistic inclinations, relying on moral values and preferences, such as equality and fairness, as well as to self-interested motivation[142]. Economic games provide an interesting way to experimentally investigate social cooperation. In the Ultimatum Game, a proposer makes an offer to a responder on how to split an amount of money. If the responder accepts, the money is split as proposed. However, if the responder rejects, both players end up with nothing. Recent functional imaging studies in the new area of NEUROECONOMICS show that the brain areas activated during these interactions include limbic/paralimbic regions (the hypothalamus and ventral striatum), the anterior prefrontal cortex (aPFC) and the superior temporal sulcus (STS)[143], which overlap with the regions involved in moral cognition (panel a). Activation of the insula, a paralimbic structure, predicted rejection of unfair offers[144], and activity in the aPFC and striatum reflected decisions to punish violators of the norm[145]. An interesting aspect of these experimental designs is that they make it possible to measure brain activation during real-time interactions among two or more individuals. Prec, precuneus.**

**Behavioural studies clearly underscore the role of culturally shaped preferences and values in social and economic interactions. For example, behaviour in experimental games might reflect differences in social cooperativeness, such as proneness to engage in collective efforts. In a study conducted in Tanzania, the more individualistic Pimbwe group made low offers in the Ultimatum Game, whereas the highly cooperative Sukuma group consistently made generous offers[146]. Such cultural differences are illustrated by the variability of proposals in the Ultimatum Game among different social groups (panel b), although the underlying cognitive and motivational mechanisms and their relationships to social norms and values are still largely unknown[105]. Future studies could address the distinct roles of PFC subregions, limbic areas and the temporal cortex in representing culturally shaped moral values and norms. Panel a reproduced, with permission, from REF. 143 © (2004) Elsevier Science. Panel b modified, with permission, from REF. 146 © (2005) Sigma Xi, The Scientific Research Society.**

evidence that PFC damage leads to universal impairments in these processes, and it is hard to imagine how complex personality and emotional changes could emerge from dysfunction of these all-purpose processes[97]. The finding that performance on social reasoning tasks crucially depends on the content of the information being evaluated (for example, social versus non-social)[90,91,94,98], and evidence from functional imaging and brain lesion studies linking PFC subregions to content-specific dissociations in social reasoning[99], ATTITUDES[82,83], beliefs[92] and emotional signals[100] indicate that a representational view can better explain the role of the PFC in moral cognition.

***Culture and the brain.*** Finally, inferring cognitive and neural mechanisms from behaviours can be misleading[101], especially when cultural and situational factors are involved. For instance, Westerners and East Asians differ in categorization strategies when making causal attributions and predictions[102], and moral values and social preferences are shaped by cultural codification[103–105]. The PFC has a central role in the internalization of moral values and norms through the integration of cultural and contextual information during development[24,106,107]. Assessing the relationships between culturally shaped values and preferences in social interactions will therefore be a logical next step in designing experiments with which to study moral cognition (BOX 2).

## A new model: EFECs

The evidence discussed above strongly indicates that the neural mechanisms of moral cognition are not restricted to the PFC, limbic areas or any other brain region. We propose a new representational neural architecture, designed to circumvent the limitations of previous frameworks. In our view, moral cognitive phenomena emerge from the integration of content- and context-dependent representations in cortical–limbic networks.

The structure of the framework, its properties and its predictions rely on three main components (FIG. 3a): structured event knowledge, which corresponds to context-dependent representations of events and event sequences in the PFC; social perceptual and functional features, represented as context-independent knowledge in the anterior and posterior temporal cortex; and central motive and emotional states, which correspond to context-independent activation in limbic and paralimbic structures. These components were derived from clinical and imaging evidence, and their relevance to moral cognition and behaviour is reviewed below. Component representations interact and give rise to event–feature–emotion complexes (EFECs) through three putative BINDING mechanisms: sequential binding, which has been proposed to link SECs in the PFC[108]; temporal binding among anatomically highly connected regions, also involved in PERCEPTUAL GESTALTS in the posterior cortex[109]; and third-party binding of anatomically loosely connected regions by synchronized activity, which results in the formation of episodic memories[108,110].

***Structured event knowledge.*** Morality is a real-world business. It is about people navigating, interacting and making choices in an ever-changing world. Humans integrate extensive contextual elements when assessing the behaviour of others and when appreciating their own actions in a given situation. The importance of the PFC in structuring context-dependent social and non-social knowledge into SECs is described in terms of the SEC framework[14]. Distinct PFC regions have been postulated to be involved in representing event sequence knowledge.

According to the SEC model, over-learned event sequences, such as routine tasks, are stored in medial and more posterior sectors of the PFC, whereas less predictable event sequences are represented in the DLPFC. The anterior sectors of the PFC are more important for storing long-term goals and multi-stage event complexes, such as those involved in making plans and thinking about the future[23,84,111–113], and have been implicated in integrating separate cognitive operations to achieve a superordinate behavioural goal[84,114]. Finally, the ventromedial sectors of the PFC are preferentially involved in representing social and emotional event knowledge, which is essential for the formation of attitudes and social stereotypes[115–117].

***Social perceptual and functional features.*** When you skim your favourite newspaper, gather at a conference or attend a family meeting, your brain deals with a massive number of perceptual signs of social significance. Our ability to manage this burden of information relies on complex patterns of featural and semantic knowledge[118]. The existence of context-independent featural representations is supported by a vast amount of neuropsychological and functional imaging evidence[119,120]. Making implicit or explicit moral appraisals when engaged in the social world requires the ability to efficiently extract social perceptual and functional features from the environment. Social perceptual features are extracted from facial expression, gaze, prosody, body posture and gestures. The posterior STS is a key region for storing these representations[33,121]. In support of this view, morphological abnormalities of the STS region have been implicated in the impaired social decoding observed in autism[122].

Social functional features code for context-independent semantic properties that are extracted from different social situations. The importance of the anterior temporal cortex for semantic feature knowledge is underscored by supramodal semantic impairments in semantic dementia[123]. Patients with anterior temporal lobe resection show impairments in naming human actions[124], which indicates that this region is involved in representing functional knowledge relevant to people. The severe behavioural changes that are associated with isolated anterior temporal atrophy in semantic dementia[74,75], and the finding of semantic impairments and abnormal activity in this brain region in psychopathic individuals[39,125] support this view.

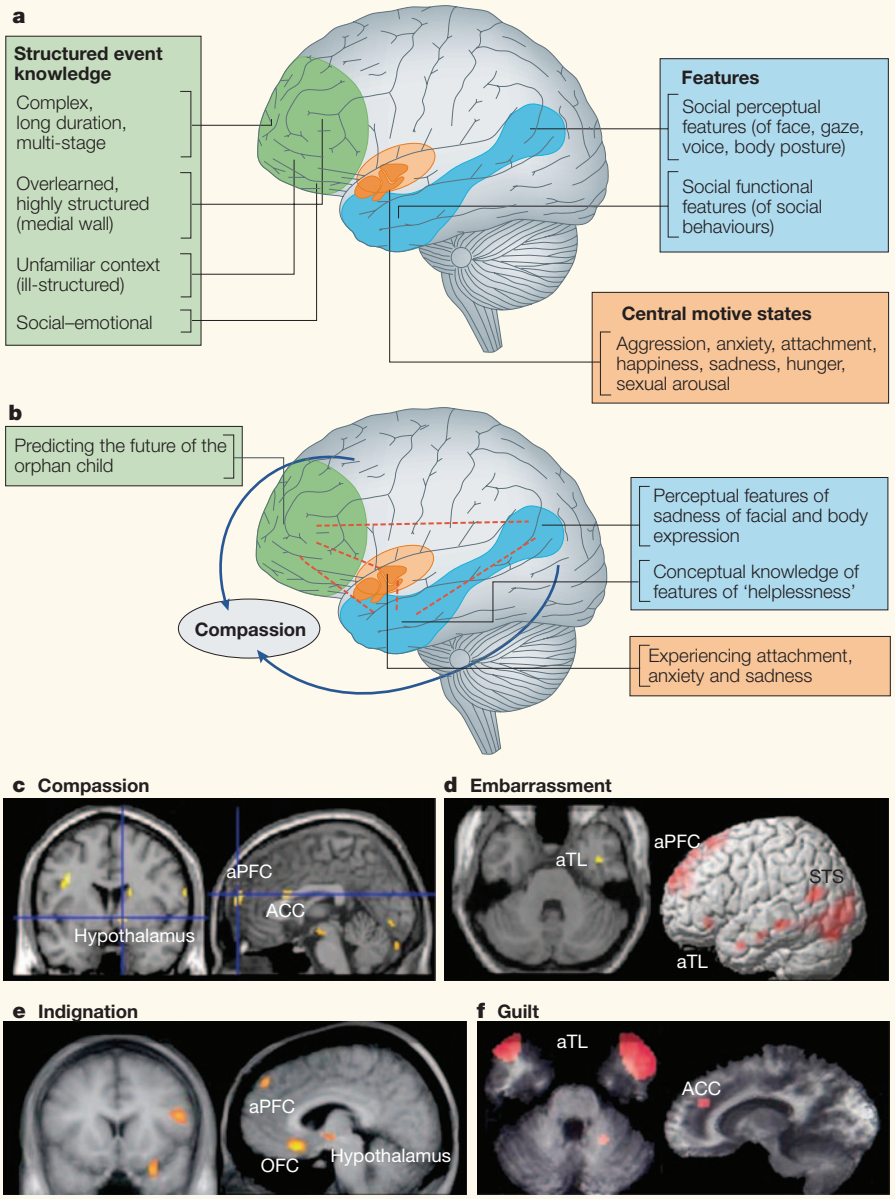



Figure 3 | **The event–feature–emotion complex framework. a** | The event–feature–emotion complex (EFEC) framework postulates that moral cognitive and behavioural phenomena arise from the binding of three main components: structured event knowledge (provided by context-dependent representations in prefrontal subregions), social perceptual and functional features (stored in the posterior and anterior sectors of the temporal cortex) and central motive or basic emotional states (such as aggressiveness, sadness, attachment or sexual arousal, represented in limbic and paralimbic regions). **b** | Emergent representations predicted by the EFEC model. Relevant types of moral cognition phenomenon that can be understood on the basis of the EFEC framework include moral emotions, moral values and long-term goals. The elements from the three main components of the EFEC framework interact to produce the moral emotion compassion. The prefrontal cortex provides contextual event representations (for example, the girl is an orphan and the odds of adoption are low), the superior temporal sulcus and anterior temporal cortex region contribute social perceptual (sad facial expression of a child) and functional (the concept of 'helplessness') features, and limbic/paralimbic regions underlie central motive states (feeling sadness, anxiety and attachment). These component representations give rise to a 'gestalt' experience by way of temporal synchronization[109]. **c–f** | Recent functional imaging studies show that these component representations are consistently activated by distinct moral emotions: compassion (Moll *et al.*, unpublished observations) (**c**); embarrassment[50] (**d**); indignation[151] (**e**); and guilt[49] (**f**). ACC, anterior cingulate cortex; aPFC, anterior prefrontal cortex; aTL, anterior temporal lobes; OFC, orbitofrontal cortex; STS, superior temporal sulcus. Anatomical image in panels **a** and **b** adapted, with permission, from REF. 148 © (1996) Appleton & Lange. Panel **d** reproduced, with permission, from REF. 50 © (2004) Elsevier Science. Panel **e** reproduced, with permission, from REF. 151 © (2005) Lippincott Williams and Wilkins. Panel **f** reproduced, with permission, from REF. 49 © (2000) Elsevier Science.

***Central motive states.*** Moral cognition depends on elaborated cortical mechanisms for representing and retrieving event knowledge, semantic information and perceptual features. However, morality would be reduced to a meaningless concept if it were stripped from its motivational and emotional aspects. Limbic and paralimbic regions[126] monitor bodily homeostasis and underlie elementary emotional or motivational 'states'. The concept of 'central motive states'[127] is an influential account of the basic mechanisms of motivation. Together with other limbic/paralimbic and brainstem structures (the amygdala, septal nuclei, ventral striatum, medial forebrain bundle, ventral tegmental area and paralimbic cortex), hypothalamic activity has a central role in 'undirected' emotionality, including sexual arousal, social attachment, hunger, aggression and extremes of pleasantness. Accordingly, these states can be potently elicited or suppressed by selective lesions, drugs and electrical stimulation of these regions, as well as by imbalances of neurotransmitters or neuromodulators[13,36,126–129]. Central motive states must be distinguished from basic emotions, such as fear and disgust. Basic emotions emerge by temporal binding of context representations (perceiving the feared object or situation) and the central motive state itself (undirected anxiety).

Several limbic nuclei exert a powerful influence over a wide range of behaviours through reciprocal connections with the PFC and other cortical regions[126,130]. Our framework underscores a key role for central motive states in moral behaviour by way of integrated cortical–limbic networks. For example, cortical representations allow you to notice that someone is hurt, whereas central motive states elicit anxiety and attachment, which encourage you to help the suffering person. This integrative perspective contrasts with the commonly held view that 'rational' cognitive mechanisms control or compete with emotional ones.

## Explaining complex moral phenomena

Although the EFEC framework can predict several possible emergent properties, we discuss three of the most relevant for moral cognition: moral emotions, moral values and long-term goals.

Whereas basic emotions spring from perceptions, imagination or recollections endowed with personal relevance, moral emotions are linked to the interest or welfare of other individuals or society as a whole[131]. Guilt, compassion, embarrassment, shame, pride, contempt and gratitude are prototypical examples of moral emotions; depending on the context, other emotions — such as disgust, awe and indignation or anger — may also qualify as moral emotions[131,132]. As a general rule, moral emotions result from interactions among values, norms and contextual elements of social situations, and are elicited in response to violations or enforcement of social preferences and expectations[104,132]. Although the contextual cues that link moral emotions to social norms are variable and shaped by culture[103], these emotions evolved from prototypes found in other primates[11] and can be characterized across cultures[133].

> "...moral emotions result from interactions among values, norms and contextual elements of social situations, and are elicited in response to violations or enforcement of social preferences and expectations."

Moral emotions require the integration of the three components of the model. For example, compassion requires the integration of context-independent social perceptual features (for instance, 'a sad facial expression of a child'), social functional features (abstract conceptual knowledge pointing to the features of 'helplessness' of an orphan child), and central motive states (sadness, anxiety and attachment) with specific contextual event representations (such as 'her parents died in an accident, and the chances of adoption at her age are low') (FIG. 3b).

Moral values (for example, being an honest citizen or a caring parent) and norms (such as paying taxes and not stealing) comprise several standards of conduct in society; they enforce social conformity and shape attitudes and expectations in social situations[105,134]. Behaviours that deviate from or enforce these values elicit different moral judgements and emotions (for instance, pride when one upholds the values, or guilt when one fails to do so). Despite the intimate link between moral values, norms and attitudes with moral cognition, their neural representations are still poorly understood. Recent functional MRI (fMRI) studies have started to shed light on these aspects. Attitudes that relate to sensitive issues, such as war, murder and abortion, activate networks involving different PFC sectors, limbic and paralimbic regions and the anterior temporal cortex[80,135]. In our view, the moral values and moral emotions involved in specific situations directly influence implicit and explicit moral appraisals.

Another key aspect of moral cognition is the representation of goals and the prediction of the utility of outcomes[136] in social situations. Pursuing goals or foreseeing possible consequences of one's decisions in the social world requires the ability to estimate the likelihood of outcomes and their desirability. Functional integration of information in the anterior PFC (which represents long-term outcomes)[88] and limbic structures (which code for the reward value of behavioural choices) is key to our ability to weigh the motivational relevance of different behavioural choices in social situations[13]. This view can be parsimoniously integrated with cognitive and neurobiological models of reward expectation and utility estimation[65,137–139], and contrasts with the interpretation that the PFC performs a 'cognitive role' in abstract moral reasoning by suppressing emotional responses[47]. Our view posits a central role for the human ability to represent and evaluate large sets of possible event outcomes, which are linked to motivational salience through cortical–limbic integration.

## Model predictions

The EFEC framework allows us to generate new predictions about the patterns of moral behavioural changes that result from dysfunction of different brain regions that cannot be made using the other frameworks described above. In addition, it offers novel ways of interpreting functional imaging findings in healthy individuals. Some of these predictions are described below.

A general prediction is that different neural subdivisions store distinct knowledge or motivational states. The binding of particular neuronal groups in each of these areas could give rise to a particular moral cognitive representation (FIG. 3b).

A lesion of the anterior PFC would lead to selective impairments in moral evaluations that rely on predicting the long-term outcomes of one's own actions, such as the anticipation of guilt. We predict that patients with damage to this area would be guided more by short-term goals because their knowledge of long-term plans and goals, or

their binding with motivational relevance is impaired. In our interpretation, the activation of this region during moral judgement results from representing possible outcomes and how they branch into the future; this offers a parsimonious explanation for anterior PFC activation in reflective moral reasoning ('moral calculus')[13], and in 'utilitarian' moral judgements[47].

Lesions of the DLPFC would lead to behavioural impairments in unfamiliar situations, in which reliance on external guidance and stimuli becomes an issue[54], but would leave intact well-established social behaviours and attitudes. By contrast, lesions of the ventral sectors of the PFC would lead to severe social behavioural changes due to disruption of social–emotional contextual knowledge[14], with early lesions having more drastic effects as they impair the learning of moral values[107]. Lesions of the ventromedial PFC would tend to impair adherence to well-established social norms and attitudes, which is consistent with the often ensuing personality changes. Lesions of the lateral OFC are expected to impair behaviours that rely on dynamically comparing non-matching social–emotional cues with stored representations, which is in agreement with the proposed role of this region in social response reversal[62].

Damage to the posterior STS is predicted to disrupt the ability to recognize socially relevant perceptual features of faces, body posture and movements. This would lead to inadequate social behaviour under circumstances that depend on the perception of these signals, but would leave intact previously established social rules, attitudes and outcome knowledge, as well as their integration with emotional and motivational states. Therefore, acquired lesions in adulthood are predicted to have a relatively minor effect on general social knowledge. However, early developmental disorders that affect this region would impair the acquisition of general social knowledge, including social rules, attitudes and outcome knowledge, which depends on the perceptual integration of social situations.

Lesions of the anterior temporal lobe are expected to disrupt knowledge of social concepts and values that are more context-independent (such as 'honour' and 'greed'), but to leave intact highly context-dependent knowledge of sequences of social events (for example, 'going to a supermarket'). We predict that loss of this knowledge, as measured by semantic memory tasks, would impair implicit and explicit evaluation of one's own and others' social behaviours.

## Glossary

ATTITUDES
Context-dependent, emotionally laden social concepts and intuitions.

BASIC EMOTIONS
A collection of emotions that are shared by most mammals (for example, fear, sadness, disgust, anger, happiness and surprise) that can readily be recognized from facial expressions (mimicry), gaze direction, voice intonation, gestures and body postures.

BINDING
Temporal synchronization of different neuronal assemblies, which correspond to stored neural representations, or codes.

EXTINCTION
The mechanism by which a previously learned automatic behavioural response is extinguished.

IOWA GAMBLING TASK
A card-sorting task designed to probe implicit mechanisms that govern individual choices in reward and punishment contexts.

MORAL EMOTIONS
Emotions that are linked to the interest or welfare of other people or society as a whole.

MORAL JUDGEMENT
A type of evaluative judgement that is based on assessments of the adequacy of one's own and others' behaviours according to socially shaped ideas of right and wrong.

MORAL REASONING
The thinking mechanism through which moral judgements are attained.

MORAL VALUES
Culturally shaped concepts and attitudes that code for personal and societal preferences and standards.

NEUROECONOMICS
An interdisciplinary field that aims to understand cognitive and neurobiological mechanisms that underlie choice behaviour and utility estimation.

PERCEPTUAL GESTALT
Simultaneous perception of sensory stimuli in one or more sensory modalities, experienced as a unified, integrated pattern.

PSYCHOPATHY
A severe form of antisocial personality disorder, characterized by callousness and lack of empathy.

RESPONSE REVERSAL
A change in a learned behavioural response following a change in reinforcement contingencies.

SECOND-ORDER FALSE BELIEF TASKS
Sophisticated mind-reading tasks that require the evaluation of what another person believes that a third person is thinking.

THEORY OF MIND
A specific cognitive ability that allows one to understand other people as intentional, perceptive and emotional agents, or to interpret their minds in terms of intentional, perceptual or feeling states.

UTILITARIANISM
A moral philosophical theory according to which the best decisions are those that lead to the higher overall degree of happiness or well-being for the greatest number of people.

Dysfunction of limbic or paralimbic regions is predicted to cause exaggeration or attenuation of basic motivational and emotional states, thereby affecting moral behaviour. Lesions of the hypothalamus, septal nuclei, basal forebrain and neighbouring structures are predicted to produce gross distortions of the valence of moral values, attitudes and moral emotions. This is in line with the observation of unprovoked rage, lack of empathy and abnormal sexual behaviours following isolated damage to limbic and paralimbic regions[35,36,128,140]. In the case of acquired lesions in adulthood, gross changes in the motivational relevance of behaviours would be observed, in spite of preserved knowledge of social rules. By contrast, early developmental disorders that affect these regions would cause aberrant social learning. Abnormal behaviours in these patients do not result from impaired inhibitory mechanisms, but from a lack of emotional empathy, or increased aggression or sexual drive, for example. These motivational states can be investigated with functional imaging and physiological methods (such as galvanic skin responses).

## Conclusions and future directions

Moral cognitive neuroscience researchers have developed innovative paradigms for the scientific exploration of unique forms of human social behaviour. Recent studies are fostering new interpretations with regard to the neural bases of moral cognition. However, they are also generating new conundrums that require theoretical frameworks to be compatible with distinctive characteristics of the human moral condition.

We have reviewed clinical and experimental work and discussed the strengths and limitations of current theoretical accounts that are relevant to moral cognitive neuroscience. We have proposed a new comprehensive model — the event–feature–emotion complex framework — which integrates cultural and context-dependent knowledge, semantic social knowledge and basic motivational states. This framework allows us to generate testable predictions for neuropsychological dissociations associated with selective brain dysfunction, and can be used as a guideline for designing future experiments.

Moral cognitive neuroscience can improve assessment, prediction and treatment of behavioural disorders. Understanding the neural basis of moral cognition will help to shape environmental, psychological and medical intervention aimed at promoting prosocial behaviours and social welfare. Future studies will be needed to explore the neural basis of how different individuals and social groups make use of strategies and heuristics to solve moral conflicts. The implications of this new knowledge for how societies conduct business, regulate social behaviour and plan for their futures remain to be seen.

*Jorge Moll, Roland Zahn, Frank Krueger and Jordan Grafman are at The Cognitive Neuroscience Section, National Institute of Neurological Disorders and Stroke, Building 10; Room 5C205; MSC 1440, NIH, Bethesda, Maryland 20892-1440, USA.*

*Ricardo de Oliveira-Souza is at the Cognitive and Behavioral Neuroscience Unit, LABS-D'Or Hospital Network, R. Pinheiro Guimaraes 22, 3rd floor, Rio de Janeiro 22281-080, Brazil.*

*Correspondence to J.G.*
*e-mail: grafmanj@ninds.nih.gov*

doi:1038/nrn1768

1. Blakemore, S.-J., Winston, J. & Frith, U. Social cognitive neuroscience: where are we heading? *Trends Cogn. Sci.* **8**, 216–222 (2004).
2. Wood, J. N. Social cognition and the prefrontal cortex. *Behav. Cogn. Neurosci. Rev.* **2**, 97–114 (2003).
3. Adolphs, R. Cognitive neuroscience of human social behaviour. *Nature Rev. Neurosci.* **4**, 165–178 (2003).
4. MacIntyre, A. *After Virtue* (Duckworth, London, 1985).
5. Tranel, D. 'Acquired sociopathy': the development of sociopathic behavior following focal brain damage. *Prog. Exp. Pers. Psychopathol. Res.* 285–311 (1994).
6. Casebeer, W. D. Moral cognition and its neural constituents. *Nature Rev. Neurosci.* **4**, 840–846 (2003).
7. Greene, J. From neural 'is' to moral 'ought': what are the moral implications of neuroscientific moral psychology? *Nature Rev. Neurosci.* **4**, 846–849 (2003).
8. Casebeer, W. D. *Natural Ethical Facts: Evolution, Connectionism, and Moral Cognition* (MIT Press, Cambridge, Massachusetts, USA, 2003).
9. Schulkin, J. *Roots of Social Sensitivity and Neural Function* (MIT Press, Cambridge, Massachusetts, USA, 2000).
10. Hauser, M. D., Chen, M. K., Chen, F. & Chuang, E. Give unto others: genetically unrelated cotton-top tamarin monkeys preferentially give food to those who altruistically give food back. *Proc. Biol. Sci.* **270**, 2363–2370 (2003).
11. de Waal, F. B. M. *Tree of Origin: What Primate Behavior Can Tell Us About Human Social Evolution* (Harvard Univ. Press, Cambridge, Massachusetts, USA, 2001).
12. Allman, J., Hakeem, A. & Watson, K. Two phylogenetic specializations in the human brain. *Neuroscientist* **8**, 335–346 (2002).
13. Moll, J., de Oliveira-Souza, R. & Eslinger, P. J. Morals and the human brain: a working model. *Neuroreport* **14**, 299–305 (2003).
14. Wood, J. N. & Grafman, J. Human prefrontal cortex: processing and representational perspectives. *Nature Rev. Neurosci.* **4**, 139–147 (2003).
15. Grafman, J. Similarities and distinctions among current models of prefrontal cortical functions. *Ann. NY Acad. Sci.* **769**, 337–368 (1995).
16. Mithen, S. *The Prehistory of the Mind: The Cognitive Origins of Art, Religion and Science* (Thames and Hudson, London, 1996).
17. Altschuler, E. L., Haroun, A., Ho, B. & Weimer, A. Did Samson have antisocial personality disorder? *Arch. Gen. Psychiatry* **58**, 202–203 (2001).
18. Augstein, H. F. J C Prichard's concept of moral insanity — a medical theory of the corruption of human nature. *Med. Hist.* **40**, 311–343 (1996).
19. Prichard, J. C. in *The Cyclopaedia of Practical Medicine* (eds Forbes, J., Tweedie, A. & Conolly, J.) 10–32; 847–875 (Sherwood, Gilbert and Piper, London, 1833–1835).
20. Welt, L. Ueber charackterveranderungen des menschen. *Dtsch Arch. Klin. Med.* **42**, 339–390 (1888).
21. Macmillan, M. *An Odd Kind of Fame: Stories of Phineas Gage* (MIT Press, Cambridge, Massachusetts, USA, 2000).
22. Grafman, J. *et al.* Frontal lobe injuries, violence, and aggression: a report of the Vietnam Head Injury Study. *Neurology* **46**, 1231–1238 (1996).
23. Eslinger, P. J. & Damasio, A. R. Severe disturbance of higher cognition after bilateral frontal lobe ablation: patient EVR. *Neurology* **35**, 1731–1741 (1985).
24. Anderson, S. W., Bechara, A., Damasio, H., Tranel, D. & Damasio, A. R. Impairment of social and moral behavior related to early damage in human prefrontal cortex. *Nature Neurosci.* **2**, 1032–1037 (1999).
25. Eslinger, P. J., Grattan, L. M., Damasio, H. & Damasio, A. R. Developmental consequences of childhood frontal lobe damage. *Arch. Neurol.* **49**, 764–769 (1992).
26. Cleckley, H. *The Mask of Sanity* (CV Mosby, St Louis, Missouri, USA, 1964).
27. Hare, R. D. *Psychopathy: Theory and Research* (John Wiley, New York, USA, 1970).
28. Miller, B. L., Chang, L., Mena, I., Boone, K. & Lesser, I. M. Progressive right frontotemporal degeneration: clinical, neuropsychological and SPECT characteristics. *Dementia* **4**, 204–213 (1993).
29. Perry, R. J. *et al.* Hemispheric dominance for emotions, empathy and social behaviour: evidence from right and left handers with frontotemporal dementia. *Neurocase* **7**, 145–160 (2001).
30. Tranel, D., Bechara, A. & Denburg, N. L. Asymmetric functional roles of right and left ventromedial prefrontal cortices in social conduct, decision-making, and emotional processing. *Cortex* **38**, 589–612 (2002).
31. Eslinger, P. J. Adolescent neuropsychological development after early right prefrontal cortex damage. *Dev. Neuropsychol.* **18**, 297–329 (2001).
32. Kruesi, M. J., Casanova, M. F., Mannheim, G. & Johnson-Bilder, A. Reduced temporal lobe volume in early onset conduct disorder. *Psychiatry Res.* **132**, 1–11 (2004).
33. Allison, T., Puce, A. & McCarthy, G. Social perception from visual cues: role of the STS region. *Trends Cogn. Sci.* **4**, 267–278 (2000).
34. Frith, C. D. & Frith, U. Interacting minds — a biological basis. *Science* **286**, 1692–1695 (1999).
35. Burns, J. M. & Swerdlow, R. H. Right orbitofrontal tumor with pedophilia symptom and constructional apraxia sign. *Arch. Neurol.* **60**, 437–440 (2003).
36. Weissenberger, A. A. *et al.* Aggression and psychiatric comorbidity in children with hypothalamic hamartomas and their unaffected siblings. *J. Am. Acad. Child Adolesc. Psychiatry* **40**, 696–703 (2001).
37. Muller, J. L. *et al.* Abnormalities in emotion processing within cortical and subcortical regions in criminal psychopaths: evidence from a functional magnetic resonance imaging study using pictures with emotional content. *Biol. Psychiatry* **54**, 152–162 (2003).
38. Soderstrom, H. *et al.* Reduced frontotemporal perfusion in psychopathic personality. *Psychiatry Res* **114**, 81–94 (2002).
39. Kiehl, K. A. *et al.* Limbic abnormalities in affective processing by criminal psychopaths as revealed by functional magnetic resonance imaging. *Biol. Psychiatry* **50**, 677–684 (2001).
40. Raine, A., Lencz, T., Bihrle, S., LaCasse, L. & Colletti, P. Reduced prefrontal gray matter volume and reduced autonomic activity in antisocial personality disorder. *Arch. Gen. Psychiatry* **57**, 119–127; discussion 128–129 (2000).
41. Beer, J. S., Heerey, E. A., Keltner, D., Scabini, D. & Knight, R. T. The regulatory function of self-conscious emotion: insights from patients with orbitofrontal damage. *J. Pers. Soc. Psychol.* **85**, 594–604 (2003).
42. Camille, N. *et al.* The involvement of the orbitofrontal cortex in the experience of regret. *Science* **304**, 1167–1170 (2004).
43. Moll, J., Eslinger, P. J. & Oliveira-Souza, R. Frontopolar and anterior temporal cortex activation in a moral judgment task: preliminary functional MRI results in normal subjects. *Arq. Neuropsiquiatr.* **59**, 657–664 (2001).
44. Moll, J., de Oliveira-Souza, R., Bramati, I. E. & Grafman, J. Functional networks in emotional moral and nonmoral social judgments. *Neuroimage* **16**, 696–703 (2002).
45. Heekeren, H. R., Wartenburger, I., Schmidt, H., Schwintowski, H. P. & Villringer, A. An fMRI study of simple ethical decision-making. *Neuroreport* **14**, 1215–1219 (2003).
46. Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M. & Cohen, J. D. An fMRI investigation of emotional engagement in moral judgment. *Science* **293**, 2105–2108 (2001).
47. Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M. & Cohen, J. D. The neural bases of cognitive conflict and control in moral judgment. *Neuron* **44**, 389–400 (2004).
48. Moll, J. *et al.* The neural correlates of moral sensitivity: a functional magnetic resonance imaging investigation of basic and moral emotions. *J. Neurosci.* **22**, 2730–2736 (2002).
49. Shin, L. M. *et al.* Activation of anterior paralimbic structures during guilt-related script-driven imagery. *Biol. Psychiatry* **48**, 43–50 (2000).
50. Takahashi, H. *et al.* Brain activation associated with evaluative processes of guilt and embarrassment: an fMRI study. *Neuroimage* **23**, 967–974 (2004).
51. Berthoz, S., Armony, J. L., Blair, R. J. & Dolan, R. J. An fMRI study of intentional and unintentional (embarrassing) violations of social norms. *Brain* **125**, 1696–1708 (2002).
52. Dougherty, D. D. *et al.* Anger in healthy men: a PET study using script-driven imagery. *Biol. Psychiatry* **46**, 466–472 (1999).
53. Heekeren, H. R. *et al.* Influence of bodily harm on neural correlates of semantic and moral decision-making. *Neuroimage* **24**, 887–897 (2005).
54. Miller, E. K. & Cohen, J. D. An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* **24**, 167–202 (2001).
55. Fellows, L. K. & Farah, M. J. Is anterior cingulate cortex necessary for cognitive control? *Brain* **128**, 788–796 (2005).
56. Damasio, A. R., Tranel, D. & Damasio, H. Individuals with sociopathic behavior caused by frontal damage fail to respond autonomically to social stimuli. *Behav. Brain Res.* **41**, 81–94 (1990).
57. Newman, J. P., Patterson, C. M. & Kosson, D. S. Response perseveration in psychopaths. *J. Abnorm. Psychol.* **96**, 145–148 (1987).
58. Bechara, A., Damasio, H., Tranel, D. & Damasio, A. R. Deciding advantageously before knowing the advantageous strategy. *Science* **275**, 1293–1295 (1997).
59. Bechara, A., Tranel, D. & Damasio, H. Characterization of the decision-making deficit of patients with ventromedial prefrontal cortex lesions. *Brain* **123**, 2189–2202 (2000).
60. Zahn, T. P., Grafman, J. & Tranel, D. Frontal lobe lesions and electrodermal activity: effects of significance. *Neuropsychologia* **37**, 1227–1241 (1999).
61. Maia, T. V. & McClelland, J. L. A reexamination of the evidence for the somatic marker hypothesis: what participants really know in the Iowa gambling task. *Proc. Natl Acad. Sci. USA* **101**, 16075–16080 (2004).
62. Blair, R. J. & Cipolotti, L. Impaired social response reversal. A case of 'acquired sociopathy'. *Brain* **123**, 1122–1141 (2000).
63. Fellows, L. K. & Farah, M. J. Different underlying impairments in decision-making following ventromedial and dorsolateral frontal lobe damage in humans. *Cereb. Cortex* **15**, 58–63 (2005).
64. Rolls, E. T., Hornak, J., Wade, D. & McGrath, J. Emotion-related learning in patients with social and emotional changes associated with frontal lobe damage. *J. Neurol. Neurosurg. Psychiatry* **57**, 1518–1524 (1994).
65. Rolls, E. T. The orbitofrontal cortex and reward. *Cereb. Cortex* **10**, 284–294 (2000).
66. Kringelbach, M. L. & Rolls, E. T. Neural correlates of rapid reversal learning in a simple model of human social interaction. *Neuroimage* **20**, 1371–1383 (2003).
67. O'Doherty, J., Kringelbach, M. L., Rolls, E. T., Hornak, J. & Andrews, C. Abstract reward and punishment representations in the human orbitofrontal cortex. *Nature Neurosci.* **4**, 95–102 (2001).
68. Hornak, J. *et al.* Reward-related reversal learning after surgical excisions in orbito-frontal or dorsolateral prefrontal cortex in humans. *J. Cogn. Neurosci.* **16**, 463–478 (2004).
69. Blair, R. J. Neurocognitive models of aggression, the antisocial personality disorders, and psychopathy. *J. Neurol. Neurosurg. Psychiatry* **71**, 727–731 (2001).
70. Blair, R. J. The roles of orbital frontal cortex in the modulation of antisocial behavior. *Brain Cogn.* **55**, 198–208 (2004).
71. Adolphs, R., Tranel, D. & Damasio, A. R. The human amygdala in social judgment. *Nature* **393**, 470–474 (1998).

72. Miller, B. L., Darby, A., Benson, D. F., Cummings, J. L. & Miller, M. H. Aggressive, socially disruptive and antisocial behaviour associated with fronto-temporal dementia. *Br. J. Psychiatry* **170**, 150–154 (1997).
73. Rankin, K. P. *et al.* Right and left medial orbitofrontal volumes show an opposite relationship to agreeableness in FTD. *Dement. Geriatr. Cogn. Disord.* **17**, 328–332 (2004).
74. Mendez, M. F., Chow, T., Ringman, J., Twitchell, G. & Hinkin, C. H. Pedophilia and temporal lobe disturbances. *J. Neuropsychiatry Clin. Neurosci.* **12**, 71–76 (2000).
75. Bozeat, S., Gregory, C. A., Ralph, M. A. & Hodges, J. R. Which neuropsychiatric and behavioural features distinguish frontal and temporal variants of frontotemporal dementia from Alzheimer's disease? *J. Neurol. Neurosurg. Psychiatry* **69**, 178–186 (2000).
76. Lough, S., Gregory, C. & Hodges, J. R. Dissociation of social cognition and executive function in frontal variant frontotemporal dementia. *Neurocase* **7**, 123–130 (2001).
77. Baron-Cohen, S. Out of sight or out of mind? Another look at deception in autism. *J. Child Psychol. Psychiatry* **33**, 1141–1155 (1992).
78. Richell, R. A. *et al.* Theory of mind and psychopathy: can psychopathic individuals read the 'language of the eyes'? *Neuropsychologia* **41**, 523–526 (2003).
79. Ruchkin, D. S., Grafman, J., Cameron, K. & Berndt, R. S. Working memory retention systems: a state of activated long-term memory. *Behav. Brain Sci.* **26**, 709–728; discussion 728–777 (2003).
80. Wood, J. N., Romero, S. G., Knutson, K. M. & Grafman, J. Representation of attitudinal knowledge: role of prefrontal cortex, amygdala and parahippocampal gyrus. *Neuropsychologia* **43**, 249–259 (2005).
81. Wood, J. N., Romero, S. G., Makale, M. & Grafman, J. Category-specific representations of social and nonsocial knowledge in the human prefrontal cortex. *J. Cogn. Neurosci.* **15**, 236–248 (2003).
82. Mah, L. W., Arnold, M. C. & Grafman, J. Deficits in social knowledge following damage to ventromedial prefrontal cortex. *J. Neuropsychiatry Clin. Neurosci.* **17**, 66–74 (2005).
83. Mah, L., Arnold, M. C. & Grafman, J. Impairment of social perception associated with lesions of the prefrontal cortex. *Am. J. Psychiatry* **161**, 1247–1255 (2004).
84. Koechlin, E., Basso, G., Pietrini, P., Panzer, S. & Grafman, J. The role of the anterior prefrontal cortex in human cognition. *Nature* **399**, 148–151 (1999).
85. Koechlin, E., Corrado, G., Pietrini, P. & Grafman, J. Dissociating the role of the medial and lateral anterior prefrontal cortex in human planning. *Proc. Natl Acad. Sci. USA* **97**, 7651–7656 (2000).
86. Knutson, K. M., Wood, J. N. & Grafman, J. Brain activation in processing temporal sequence: an fMRI study. *Neuroimage* **23**, 1299–1307 (2004).
87. Wood, J. N., Knutson, K. M. & Grafman, J. Psychological structure and neural correlates of event knowledge. *Cereb. Cortex* (2004).
88. Tanaka, S. C. *et al.* Prediction of immediate and future rewards differentially recruits cortico-basal ganglia loops. *Nature Neurosci.* **7**, 887–893 (2004).
89. Thomson, J. J. *Rights, Restitution, and Risk: Essays, in Moral Theory* (Harvard Univ. Press, Cambridge, Massachusetts, USA, 1986).
90. Fiddick, L., Cosmides, L. & Tooby, J. No interpretation without representation: the role of domain-specific representations and inferences in the Wason selection task. *Cognition* **77**, 1–79 (2000).
91. Cosmides, L. & Tooby, J. Dissecting the computational architecture of social inference mechanisms. *Ciba Found. Symp.* **208**, 132–156; discussion 156–161 (1997).
92. Goel, V. & Dolan, R. J. Explaining modulation of reasoning by belief. *Cognition* **87**, B11–B22 (2003).
93. Goel, V. & Dolan, R. J. Reciprocal neural response within lateral and ventral medial prefrontal cortex during hot and cold reasoning. *Neuroimage* **20**, 2314–2321 (2003).
94. Wason, P. C. Reasoning about a rule. *Q. J. Exp. Psychol.* **20**, 273–281 (1968).
95. Fuster, J. M. *The Prefrontal Cortex: Anatomy, Physiology, and Neuropsychology of the Frontal Lobe* (Raven, New York, USA, 1997).
96. MacLean, P. *A Triune Concept of the Brain and Behaviour: Hincks Memorial Lecture* (Univ. Toronto Press, Toronto, Canada, 1973).
97. Eslinger, P. J. & Geder, L. in *Behavior and Mood Disorders in Focal Frontal Lobe Lesions* (eds Bogousslavsky, J. & Cummings, J. L.) 217–260 (Cambridge Univ. Press, Cambridge, Massachusetts, USA, 2000).
98. Fiddick, L. Domains of deontic reasoning: resolving the discrepancy between the cognitive and moral reasoning literatures. *Q. J. Exp. Psychol. A* **57**, 447–474 (2004).
99. Stone, V. E., Cosmides, L., Tooby, J., Kroll, N. & Knight, R. T. Selective impairment of reasoning about social exchange in a patient with bilateral limbic system damage. *Proc. Natl Acad. Sci. USA* **99**, 11531–11536 (2002).
100. Hornak, J. *et al.* Changes in emotion after circumscribed surgical lesions of the orbitofrontal and cingulate cortices. *Brain* **126**, 1691–1712 (2003).
101. Wilkinson, D. & Halligan, P. The relevance of behavioural measures for functional-imaging studies of cognition. *Nature Rev. Neurosci.* **5**, 67–73 (2004).
102. Nisbett, R. E. & Masuda, T. Culture and point of view. *Proc. Natl Acad. Sci. USA* **100**, 11163–11170 (2003).
103. Ehrlich, P. R. *Human Natures: Genes, Cultures, and the Human Prospect* (Island, Washington DC, USA, 2000).
104. Nichols, S. Norms with feeling: towards a psychological account of moral judgment. *Cognition* **84**, 221–236 (2002).
105. Fehr, E. & Fischbacher, U. Social norms and human cooperation. *Trends Cogn. Sci.* **8**, 185–190 (2004).
106. Grattan, L. M. & Eslinger, P. J. Long-term psychological consequences of childhood frontal lobe lesion in patient DT. *Brain Cogn.* **20**, 185–195 (1992).
107. Eslinger, P. J., Flaherty-Craig, C. V. & Benton, A. L. Developmental outcomes after early prefrontal cortex damage. *Brain Cogn.* **55**, 84–103 (2004).
108. Weingartner, H., Grafman, J., Boutelle, W., Kaye, W. & Martin, P. R. Forms of memory failure. *Science* **221**, 380–382 (1983).
109. Singer, W. Consciousness and the binding problem. *Ann. NY Acad. Sci.* **929**, 123–146 (2001).
110. O'Reilly, R. C. & Rudy, J. W. Computational principles of learning in the neocortex and hippocampus. *Hippocampus* **10**, 389–397 (2000).
111. Okuda, J. *et al.* Thinking of the future and past: the roles of the frontal pole and the medial temporal lobes. *Neuroimage* **19**, 1369–1380 (2003).
112. Eslinger, P. J. & Grattan, L. M. Altered serial position learning after frontal lobe lesion. *Neuropsychologia* **32**, 729–739 (1994).
113. Goel, V., Grafman, J., Tajik, J., Gana, S. & Danto, D. A study of the performance of patients with frontal lobe lesions in a financial planning task. *Brain* **120**, 1805–1822 (1997).
114. Ramnani, N. & Owen, A. M. Anterior prefrontal cortex: insights into function from anatomy and neuroimaging. *Nature Rev. Neurosci.* **5**, 184–194 (2004).
115. Milne, E. & Grafman, J. Ventromedial prefrontal cortex lesions in humans eliminate implicit gender stereotyping. *J. Neurosci.* **21**, RC150 (2001).
116. Pietrini, P., Guazzelli, M., Basso, G., Jaffe, K. & Grafman, J. Neural correlates of imaginal aggressive behavior assessed by positron emission tomography in healthy subjects. *Am. J. Psychiatry* **157**, 1772–1781 (2000).
117. Cunningham, W. A., Raye, C. L. & Johnson, M. K. Implicit and explicit evaluation: fMRI correlates of valence, emotional intensity, and control in the processing of attitudes. *J. Cogn. Neurosci.* **16**, 1717–1729 (2004).
118. McClelland, J. L. & Rogers, T. T. The parallel distributed processing approach to semantic cognition. *Nature Rev. Neurosci.* **4**, 310–322 (2003).
119. Martin, A. & Chao, L. L. Semantic memory and the brain: structure and processes. *Curr. Opin. Neurobiol.* **11**, 194–201 (2001).
120. Caramazza, A. & Mahon, B. Z. The organization of conceptual knowledge: the evidence from category-specific semantic deficits. *Trends Cogn. Sci.* **7**, 354–361 (2003).
121. Frith, U. Mind blindness and the brain in autism. *Neuron* **32**, 969–979 (2001).
122. Boddaert, N. *et al.* Superior temporal sulcus anatomical abnormalities in childhood autism: a voxel-based morphometry MRI study. *Neuroimage* **23**, 364–369 (2004).
123. Hodges, J. R., Bozeat, S., Lambon Ralph, M. A., Patterson, K. & Spatt, J. The role of conceptual knowledge in object use evidence from semantic dementia. *Brain* **123**, 1913–1925 (2000).
124. Lu, L. H. *et al.* Category-specific naming deficits for objects and actions: semantic attribute and grammatical role hypotheses. *Neuropsychologia* **40**, 1608–1621 (2002).
125. Kiehl, K. A. *et al.* Temporal lobe abnormalities in semantic processing by criminal psychopaths as revealed by functional magnetic resonance imaging. *Psychiatry Res.* **130**, 27–42 (2004).
126. Saper, C. B. Hypothalamic connections with the cerebral cortex. *Prog. Brain Res.* **126**, 39–48 (2000).
127. Stellar, E. The physiology of motivation. 1954. *Psychol. Rev.* **101**, 301–311 (1994).
128. Haugh, R. M. & Markesbery, W. R. Hypothalamic astrocytoma. Syndrome of hyperphagia, obesity, and disturbances of behavior and endocrine and autonomic function. *Arch. Neurol.* **40**, 560–563 (1983).
129. Bejjani, B. P. *et al.* Aggressive behavior induced by intraoperative stimulation in the triangle of Sano. *Neurology* **59**, 1425–1427 (2002).
130. Brodal, P. *The Central Nervous System: Structure and Function* (Oxford Univ. Press, New York, USA, 2003).
131. Haidt, J. in *Handbook of Affective Sciences* (eds Davidson, R. J., Scherer, K. R. & Goldsmith, H. H.) 852–870 (Oxford Univ. Press, Oxford, USA, 2003).
132. Tangney, J. P. in *Self and Motivation: Emerging Psychological Perspectives* (eds Tesser, A., Stapel, D. A. & Wood, J. V.) 97–117 (American Psychological Association, Washington DC, USA, 2002).
133. Fessler, D. in *Beyond Nature or Nurture: Biocultural Approaches to the Emotions* (ed. Hinton, A.) 75–116 (Cambridge Univ. Press, New York, USA, 1999).
134. Haidt, J. The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychol. Rev.* **108**, 814–834 (2001).
135. Cunningham, W. A., Nezlek, J. B. & Banaji, M. R. Implicit and explicit ethnocentrism: revisiting the ideologies of prejudice. *Pers. Soc. Psychol. Bull.* **30**, 1332–1346 (2004).
136. Schultz, W. Getting formal with dopamine and reward. *Neuron* **36**, 241–263 (2002).
137. McClure, S. M., Daw, N. D. & Montague, P. R. A computational substrate for incentive salience. *Trends Neurosci.* **26**, 423–428 (2003).
138. Schultz, W. & Dickinson, A. Neuronal coding of prediction errors. *Annu. Rev. Neurosci.* **23**, 473–500 (2000).
139. Kahneman, D. & Tversky, A. Prospect theory: an analysis of decision under risk. *Econometrica* **47**, 263–291 (1979).
140. Fukatsu, R., Fujii, T., Yamadori, A., Nagasawa, H. & Sakurai, Y. Persisting childish behavior after bilateral thalamic infarcts. *Eur. Neurol.* **37**, 230–235 (1997).
141. Murphy, J. M. Psychiatric labeling in cross-cultural perspective. *Science* **191**, 1019–1028 (1976).
142. Henrich, J. *et al.* (eds) *Foundations of Human Sociality* (Oxford Univ. Press, London, UK, 2004).
143. Rilling, J. K., Sanfey, A. G., Aronson, J. A., Nystrom, L. E. & Cohen, J. D. The neural correlates of theory of mind within interpersonal interactions. *Neuroimage* **22**, 1694–1703 (2004).
144. Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E. & Cohen, J. D. The neural basis of economic decision-making in the Ultimatum Game. *Science* **300**, 1755–1758 (2003).
145. de Quervain, D. J. *et al.* The neural basis of altruistic punishment. *Science* **305**, 1254–1258 (2004).
146. Paciotti, B., Hadley, C., Holmes, C. & Mulder, M. B. Grass-roots justice in Tanzania: cultural evolution and game theory help to explain how a history of cooperation influences the success of social organizations. *Am. Scientist* **93**, 58–65 (2005).
147. *University of Iowa's Virtual Hospital* [online] <http://www.vh.org/> (2005).
148. Martin, J. H. *Neuroanatomy: Text and Atlas* 2nd edn (Appleton & Lange, Stamford, Connecticut, 1996).
149. Lang, P. J., Bradley, M. M. & Cuthbert, B. N. International affective picture system (IAPS): digitized photographs, instruction manual and affective ratings. Technical Report A-6. (Univ. Florida, Gainesville, Florida, USA, 2005).
150. Phan, K. L., Wager, T., Taylor, S. F. & Liberzon, I. Functional neuroanatomy of emotion: a meta-analysis of emotion activation studies in PET and fMRI. *Neuroimage* **16**, 331–348 (2002).
151. Moll, J. *et al.* The moral affiliations of disgust: a functional MRI study. *Cogn. Behav. Neurol.* **18**, 68–78 (2005).



## Acknowledgements

This research was partially supported by the LABS-D'Or Hospital Network and by the Intramural Research Program of the National Institute of Neurological Disorders and Stroke, National Institutes of Health.


## Competing interests statement

The authors declare no competing financial interests.

## Online links

**FURTHER INFORMATION**

**Cognitive Neuroscience Section, NINDS, NIH:** http://intra.ninds.nih.gov/Lab.asp?Org_ID=83

**Cognitive and Behavioural Neuroscience Unit, LABS-D'Or Hospital Network:** http://www.rededor.com.br/cbnu/

**Access to this interactive links box is free online.**