Advanced Review

An integrative cognitive neuroscience theory of social reasoning and moral judgment

Aron K. Barbey[1,2] and Jordan Grafman[1*]

Cognitive neuroscience has made considerable progress in understanding the involvement of the prefrontal cortex (PFC) in social cognition and moral judgment. Accumulating evidence suggests that representations within the lateral PFC enable people to orchestrate their thoughts and actions in concert with their intentions to support goal-directed social behavior. Despite the pivotal role of this region in guiding social interactions, remarkably little is known about the functional organization and forms of social knowledge mediated by the lateral PFC. Here, we review recent theoretical developments in evolutionary psychology and emerging evidence from the social and decision neuroscience literatures demonstrating the importance of the lateral PFC for orchestrating behavior on the basis of evolutionarily adaptive social norms for obligatory, prohibited, and permissible courses of action. © 2010 John Wiley & Sons, Ltd. *WIREs Cogn Sci*

An enduring question in cognitive science concerns the principles of coherence that underlie our beliefs: how do we decide that an observed course of action is wrong and why do we care so strongly about the social behavior of others? Here, we introduce an integrative cognitive neuroscience framework for understanding the cognitive and neural architecture of social reasoning and moral judgment,[1–4] drawing upon recent theoretical developments in evolutionary psychology and emerging neuroscience evidence demonstrating the central role of the prefrontal cortex (PFC) for these aspects of social cognition.

The neuroscientific study of social cognition reflects the interdisciplinary nature of modern science, with investigators from diverse academic disciplines (including anthropology, evolutionary psychology, social psychology, political science, behavioral economics, and decision neuroscience) exploring the unique social nature of human experience through a multifaceted lens (for recent reviews from the emerging field of social cognitive neuroscience, see Refs 5–7). This interdisciplinary enterprise has made considerable progress in understanding the involvement of the PFC in social cognition.[4,8–10] Accumulating evidence suggests that representations within the lateral PFC enable people to orchestrate their thoughts and actions in concert with their intentions to support social reasoning and moral judgment.[11–22] Despite the pivotal role of this region in guiding social interactions, fundamental questions remain concerning the functional organization and forms of social knowledge represented within the lateral PFC. We develop an integrative cognitive neuroscience framework for understanding the social functions of the lateral PFC, reviewing recent theoretical insights from evolutionary psychology and emerging neuroscience evidence to support the importance of this region for orchestrating social behavior on the basis of evolutionarily adaptive social norms.

We begin by reviewing the evolutionary foundations of normative social behavior, surveying contemporary research and theory from evolutionary psychology to suggest that widely shared norms of social exchange are the product of evolutionarily adaptive cognitive mechanisms. We then review the biology, evolution and ontogeny of the human PFC, and introduce a cognitive neuroscience framework for social reasoning based on evolutionarily adaptive social norms represented within the lateral PFC. Our review examines a broad range of evidence from the social and decision neuroscience literatures

*Correspondence to: grafmanj@ninds.nih.gov

[1] Cognitive Neuroscience Section, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD, 20892, USA

[2] Department of Psychology, Georgetown University, Washington, DC, 20057, USA

DOI: 10.1002/wcs.84

demonstrating that social norms for obligatory, prohibited, and permissible behavior are mediated by functionally specialized regions of the lateral PFC. We illustrate how this framework supports the integration and synthesis of a diverse body of neuroscience evidence and we draw conclusions about the role of the lateral PFC in social cognition more broadly, contributing to social knowledge networks by representing widely shared norms of social behavior and providing the foundations for moral systems of value and belief.

EVOLUTIONARY FOUNDATIONS OF NORMATIVE SOCIAL BEHAVIOR

Evolutionary psychology has made significant progress in understanding the evolutionary origins of normative social behavior, establishing the central role of social exchange in the formation of cooperative human societies. Social exchange promotes the survival of individuals who cooperate for mutual benefit—one providing a benefit to another, conditional on the recipient's providing a benefit in return. From our earliest ancestors to present day, social exchange has facilitated access to sustenance, protection and mates, and enabled people to live healthier and longer lives.[23,24] Social exchange interactions are therefore an important and recurrent human activity occurring over a sufficiently long time period for natural selection to have produced specialized cognitive and neural adaptations.[25,26] Evolutionary psychologists have proposed that social exchange operates on the basis of cognitive mechanisms that are designed to promote the survival of our species, representing normative social behavior that develops in all healthy humans and is mediated by evolutionarily adaptive neural systems.[27–40]

An empirical case for this proposal has been established on the basis of behavioral and neuroscience research elucidating the role of evolutionary design features in shaping cognitive and neural mechanisms for social exchange.[28–36] Game-theoretic models predict that for social exchange to persist within a species, members of the species must detect cheaters (i.e., individuals who do not reciprocate) and direct future benefits to reciprocators rather than cheaters.[37,38] Accumulating evidence supports this proposal, demonstrating that the mind embodies functionally specialized cognitive mechanisms for detecting cheaters[28–36] that operate according to behavior-guiding principles in the form of a conditional rule: If X provides a requested benefit to Y, then Y will provide a rationed benefit to X. A conditional rule expressing this kind of agreement to cooperate is referred to as a *social contract* and represents a normative standard for social behavior (e.g., the normative belief that mutual cooperation is obligatory and cheating prohibited).

A primary method for investigating conditional reasoning about social contracts is Wason's four-card selection task.[41–43] In the classic version of this task, participants are shown a set of four cards, placed on a table, each of which has a number on one side and a colored patch on the other. The visible faces of the cards show a 3, 8, red, and brown. Participants are then asked which card should be turned over to test the truth of the conditional rule 'If a card shows an even number on one face, then its opposite face shows a primary color (red, green, or blue)'. Conditional rules representing abstract or descriptive content typically elicit a correct response from only 5–30% of participants tested (8 and brown). This finding has been observed even when the rules tested are familiar or when participants are taught logic or given incentives.[28–34,44] In contrast, when the conditional rule expresses a social contract and a violation represents cheating (e.g. 'If she drinks beer then she is 21 years or older'), 65–80% of participants generate the correct response (she drinks beer and is not 21 years or older).[28–35] Cognitive experiments have demonstrated that this improved level of performance is sensitively regulated by the series of variables expected if this was a system optimally designed to reason about obligatory and prohibited forms of social behavior, rather than to support a broader class of inferences.[28–36,40,44]

Social contracts therefore represent behavior-guiding principles for evolutionarily adaptive forms of social exchange and are critical for drawing inferences about necessary courses of action concerning socially obligatory or prohibited behavior. From an evolutionary perspective, normative standards for necessary forms of social exchange can be distinguished from a broader class of inferences concerning possible or permissible courses of action. Social norms for (1) necessary behavior are central for the organization of society, representing strictly enforced rules for cooperation, the division of labor, and the distribution of resources. In contrast, social norms for (2) permissible behavior are critical for achieving adaptive goals within society, representing non-punishable courses of action that enable individuals to explore opportunities for reward and gain access to available resources.[27–40,45,46] We propose that evolutionary adaptations for reasoning about necessary (obligatory or prohibited) versus possible (permissible) courses of action have therefore fundamentally shaped the architecture of the mind, producing functionally distinct cognitive and neural mechanisms for reasoning about necessary and possible states of affairs. Although cognitive and neural mechanisms for these forms of inference emerged from goal-directed social behavior,[27–40] non-social inferences are also shaped by these systems, relying upon an evolutionarily adaptive neural architecture that distinguishes between these two fundamental classes of inference. Thus, we propose that evolutionarily adaptive mechanisms for social reasoning enabled domain-general representations for understanding necessity and possibility.

We examine the contributions of the human PFC to social reasoning in the following section, introducing a cognitive neuroscience framework for understanding the inferential architecture of the lateral PFC.

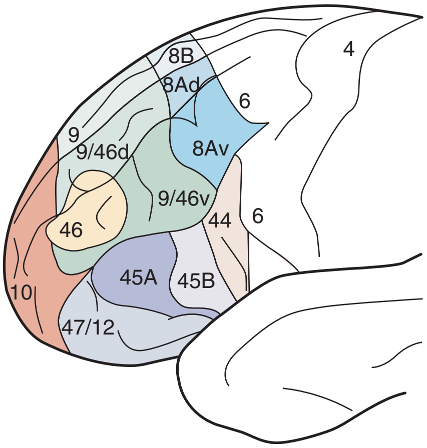


FIGURE 1 | Brodmann map of the lateral prefrontal cortex (Reprinted with permission from Ref 53 Copyright 2004 Nature Reviews).

SIMULATION THEORY OF PREFRONTAL CORTEX FUNCTION

An emerging body of evidence suggests that goal-directed social behavior centrally depends on the PFC, which is particularly important for grouping specific experiences of our interactions with the environment along common themes, that is, as behavior-guiding principles. To this end, our brains have evolved mechanisms for detecting and storing complex relationships between situations, actions, and consequences. By gleaning this knowledge from past experiences, we can develop behavior-guiding principles that allow us to infer which goals are available in similar situations in the future and what actions are likely to bring us closer to them.

Accumulating evidence demonstrates that behavior-guiding principles for social inference operate on the basis of a broadly distributed, hierarchically organized neural architecture.[47] It is widely known that experience in the physical and social world activates feature detectors in relevant feature maps of the brain (for a review on feature maps in vision, see Ref 48). When a pattern becomes active in a feature map during perception or action, conjunctive neurons in an association area capture the pattern for later cognitive use.[47,49–51]

We propose that behavior-guiding principles for social inference are mediated by higher order association areas localized within the lateral PFC. Decades of neuroscience research have demonstrated that behaviorally informative associations are encoded by the lateral PFC (for a review, see Ref 52). This work demonstrates that the lateral PFC is a site of convergence for the synthesis of multimodal information from a wide range of brain systems. The lateral PFC consists of three major subregions that each emphasizes processing of particular information based on their interconnections with specific cortical regions (Figure 1).

Ventrolateral areas are more heavily interconnected with cortical regions for processing information about visual form and stimulus identity (inferior temporal cortex), supporting the categorization of environmental stimuli in the service of goal-directed behavior. Dorsal portions of the lateral PFC are heavily interconnected with cortical areas for processing auditory, visuospatial, and motor information, enabling the regulation and control of responses to environmental stimuli. Finally, the anterolateral PFC is indirectly connected (via the ventromedial PFC) with limbic structures that process internal information, such as emotion, memory, and reward.[54–57] Together, lateral PFC subregions mediate essential elements of the external and internal environment, enabling goal-directed behavior.

Once modality-specific representations within this broadly distributed network are captured by a set of conjunctive neurons in the lateral PFC, the set can later activate the pattern in the absence of bottom-up stimulation, producing a simulation of the event sequence.[47,49–51] For example, on entering a familiar situation and recognizing it, a simulation that represents the situation becomes active. Typically not all of the situation is perceived initially. A relevant person, setting, or event may be perceived, which then suggests that a particular situation is about to play out. The simulation can be viewed as a complex configuration of multimodal components that represent the (1) situation (including agents, objects, actions, mental states, and background settings) and

(2) causal and associative relations that hold among its elements. Because part of this pattern matched the current situation initially, the larger pattern became active in memory. The remaining parts of the pattern—not yet observed in the situation—constitute inferences, namely predictions about what will occur next or explanations of observed behavior.

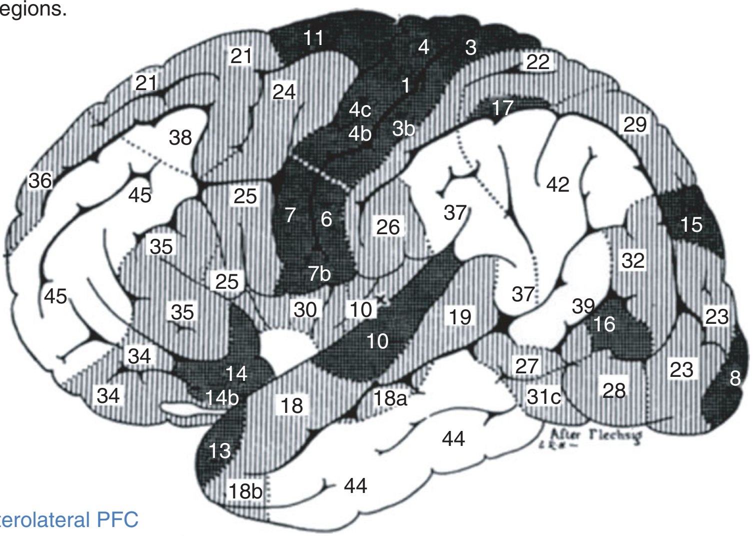
To the extent that the simulation is entrenched in memory, pattern completion is likely to occur automatically. As a situation is experienced repeatedly, its simulated components and the associations linking them increase in potency. Thus when one component is perceived initially, these strong associations complete the pattern automatically. Social norms of behavior represent deeply entrenched simulations, whose learned associations are the product of evolutionarily adaptive cognitive and neural mechanisms. The observed role of simulation mechanisms for social inference in non-human primates supports this account,[58] suggesting that modality-specific simulations represent continuity of social information processing across the species.[59] Thus, our evolutionary ancestors may have represented the social world by simulating modality-specific components of experience, providing the foundations for social processing in humans.

According to this framework, social interactions initially match modality-specific representations in one or more simulations that have become entrenched in memory. Once one of these wins the activation process, it provides inferences via pattern completion.[60] Simulations representing necessary (obligatory or prohibited) courses of action motivate expectations concerning specific actions the perceiver and recipient 'must' take, whereas simulations for possible (permissible) forms of behavior represent a broader range of outcomes, motivating expectations about courses of action the perceiver and recipient 'may' take. The unfolding of inferences about necessary and possible states of affairs—realized as a simulation—represents behavior-guiding principles for the orchestration of social thought and action. The recruitment of specific lateral PFC subregions for social inference is determined by the evolution, development, hierarchical structure, and anatomical connectivity of the PFC.

Research investigating the evolution and ontogeny of the PFC suggests that the lateral PFC initially emerged from ventrolateral prefrontal regions, followed by dorsolateral, and then anterolateral cortices[61–63] (Figure 2). From an evolutionary perspective, the emergence of lateral PFC subregions reflects their relative priority for the formation of

Dorsolateral PFC
The myelination of dorsolateral PFC subregions suggests that this area emerges after early ventrolateral PFC subregions.

Anterolateral PFC
The late myelination of the anterolateral PFC suggests that this area emerges after ventrolateral and dorsolateral profrontal regions.



Venterolateral PFC
The early myelination of ventrolateral PFC subregions suggests that this region is one of the first prefrontal areas to emerge during development.

FIGURE 2 | Ontogenetic map of the prefrontal cortex according to Flechsig.[62,63] The numeration of the areas indicates the order of their myelination. Modified with permission from Flechsig.[63]

organized social groups, with the vlPFC signaling the onset of social norms for necessary (obligatory or prohibited) courses of action, providing the foundations for standards of conduct that are central for the organization of society. Social norms for permissible behavior later enabled the representation of a broader range of possible outcomes, supporting the assessment of alternative forms of goal-directed behavior within the dlPFC. Finally, the evolution of the alPFC enabled processing of higher order relations and reasoning about complex forms of social behavior involving necessary and possible courses of action. Consistent with its evolutionary development, the ontogeny of the lateral PFC reflects the importance of first representing social norms for necessary behavior (i.e., fundamental rules the child must obey), followed by an understanding of permissible courses of action (e.g., guided by judgments of equity and fairness), and finally high-order inferences involving both forms of representation.[64]

An emerging body of evidence further demonstrates that the anterior-to-posterior axis of the lateral PFC is organized hierarchically, whereby progressively anterior subregions are associated with higher order processing requirements for planning and the selection of action (for recent reviews, see Refs 53,65–67). Thus, processes within the lateral PFC respect the hierarchical organization of this region, with progressively anterior regions representing simulations that support higher order inferences incorporating both necessary and possible states of affairs.

The connectivity of lateral PFC subregions embodies an evolutionarily adaptive neural network for goal-directed social behavior. From an evolutionary perspective, behavior requested by members of high social status represents necessary courses of action that a lower ranking individual must follow. This provides one explanation for why neural systems for identifying the social status of individuals (based on representations of visual form and stimulus identity) are anatomically connected with ventrolateral prefrontal regions for drawing inferences about necessary courses of action. An intriguing study by Marsh et al.[68] supports this proposal, demonstrating that vlPFC (area 47) is selectively recruited when processing status poses for individuals of high (rather than low) social status—providing a unified neural architecture for identifying individuals of high social status and the necessity of obeying their commands. An evolutionary perspective further suggests that social norms for possible (permissible) behavior are central for achieving adaptive goals within society,[27–40] providing one explanation for why dorsolateral prefrontal regions for drawing this type of inference are anatomically connected with brain regions for the regulation and control of behavior. Finally, adaptive behavior guided by both categories of inference draws upon higher order representations that incorporate multiple forms of social inference and therefore recruits alPFC regions that enable representations of greater complexity (e.g., incorporating emotion and memory).

We now turn to a review of emerging neuroscience evidence investigating the proposed inferential architecture of the lateral PFC.

INFERENTIAL ARCHITECTURE OF THE LATERAL PREFRONTAL CORTEX

We review a broad range of evidence from the social and decision neuroscience literatures demonstrating (1) the involvement of the vlPFC when reasoning about necessary (obligatory or prohibited) courses of action, (2) the recruitment of the dlPFC for drawing inferences about possible (permissible) states of affairs, and (3) activation in the alPFC for higher order inferences that incorporate both categories of knowledge (Figure 3). The simulation architecture underlying these forms of inference further predicts the recruitment of broadly distributed neural systems, incorporating medial prefrontal and posterior knowledge networks representing modality-specific components of experience.

Ventrolateral Prefrontal Cortex

An increasing number of social neuroscience studies have shown that social norms for necessary (obligatory or prohibited) courses of action are represented by the vlPFC (areas 44, 45, and 47; Figure 3b). Fiddick et al.[11] observed activity within bilateral vlPFC (area 47) for social exchange reasoning, employing stimuli consisting primarily of social norms for obligatory and prohibited courses of action. Converging evidence is provided by Berthoz et al.,[12] who demonstrated recruitment of left vlPFC (area 47) when participants detected violations of social norm stories representing obligatory and prohibited courses of action (e.g., the decision to 'spit out food made by the host'). Similarly, Rilling et al.[13] reported activation within left vlPFC (area 47) when participants detected the violation of obligatory and prohibited norms of social exchange in a Prisoner's dilemma game (i.e., the failure to cooperate).

The decision neuroscience literature further supports this framework, demonstrating the involvement of the vlPFC when drawing conclusions that necessarily follow from the truth of the premises, that is,

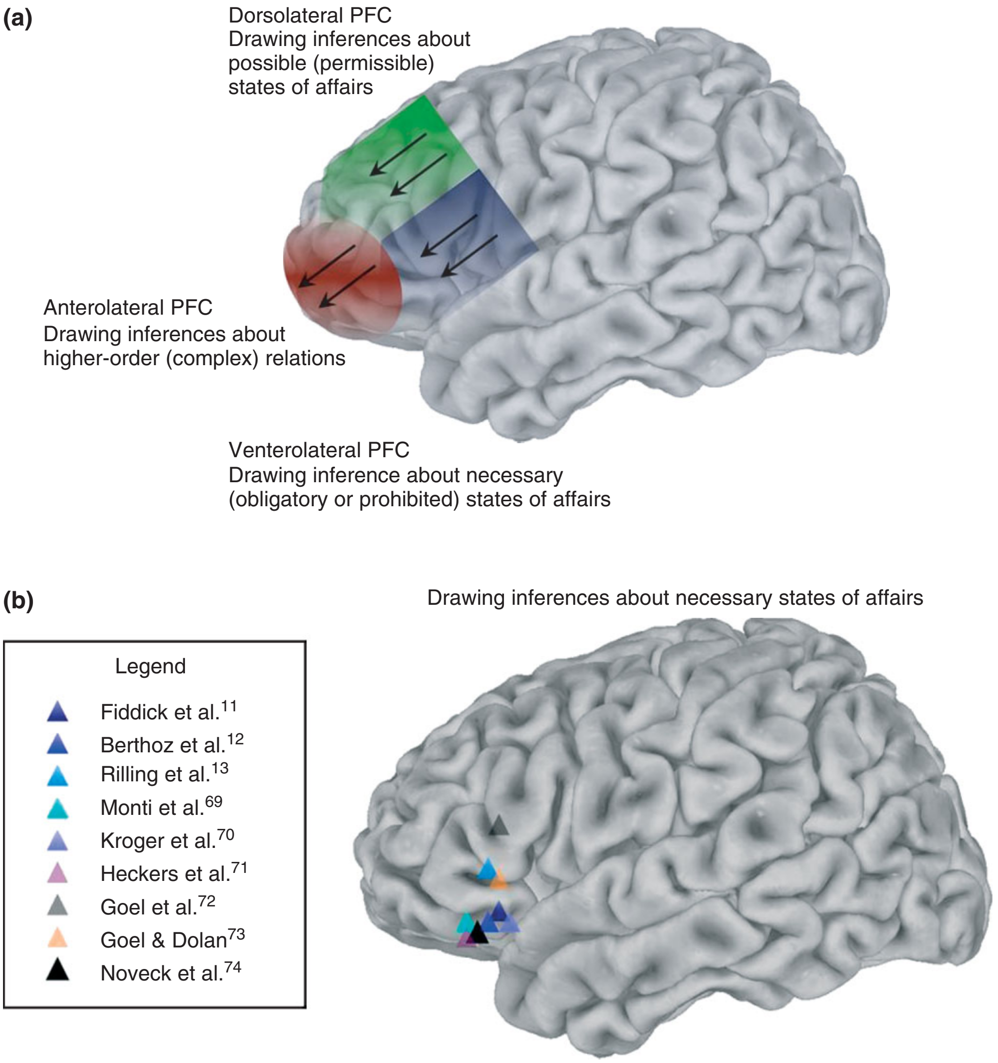

(a)
Dorsolateral PFC
Drawing inferences about possible (permissible) states of affairs
Anterolateral PFC
Drawing inferences about higher-order (complex) relations
Venterolateral PFC
Drawing inference about necessary (obligatory or prohibited) states of affairs
(b)
Drawing inferences about necessary states of affairs
Legend
▲ Fiddick et al.[11]
▲ Berthoz et al.[12]
▲ Rilling et al.[13]
▲ Monti et al.[69]
▲ Kroger et al.[70]
▲ Heckers et al.[71]
▲ Goel et al.[72]
▲ Goel & Dolan[73]
▲ Noveck et al.[74]

FIGURE 3 | An evolutionarily adaptive neural architecture for goal-directed social behavior. (a) Summarizes the functional organization of the lateral PFC, and (b–d) illustrate supportive evidence.

for *deductive inference*. Although wide consensus in the literature has not yet been reached, an increasing number of studies report consistent findings when common sources of variability are controlled (regarding the linguistic content, linguistic complexity, and deductive complexity of reasoning problems). A recent series of experiments by Monti et al.[69] controlled for these sources of variability and provided evidence that the left vlPFC (area 47) mediates representations of the logical structure of a deductive argument (e.g., If *P* or *Q*, then *Not-R*/*P*/Therefore, *Not-R*), supporting the representation of behavior-guiding principles for necessary forms of behavior within this region. Furthermore, a recent study by Kroger et al.[70] controlled for the complexity and type of calculations that were performed and also observed activation within the left vlPFC (areas 44 and 45) for deductive reasoning (see also Ref 71). Additional supporting data are provided by Goel et al.[72,73] who have consistently observed activation within the left vlPFC (areas 44 and 45) for deductive conclusions drawn from categorical syllogisms (e.g., All humans are mortal/Some animals are human/Therefore, some animals are mortal). Finally, Noveck et al.[74] demonstrated recruitment of left vlPFC (area 47) for drawing deductive conclusions from conditional statements (e.g., If *P* then *Q*/*P*/ Therefore, *Q*), consistent with the role of this region for representing behavior-guiding principles in the form of a conditional.

Dorsolateral Prefrontal Cortex

Accumulating evidence demonstrates that the dlPFC (areas 46 and 9) represents behavior-guiding principles for evaluating the permissibility or fairness of observed behavior (Figure 3c). An early study by Sanfey et al.[75] reported activity within the right dlPFC (area 46) when participants evaluated the fairness of an offer in an ultimatum game. Knoch

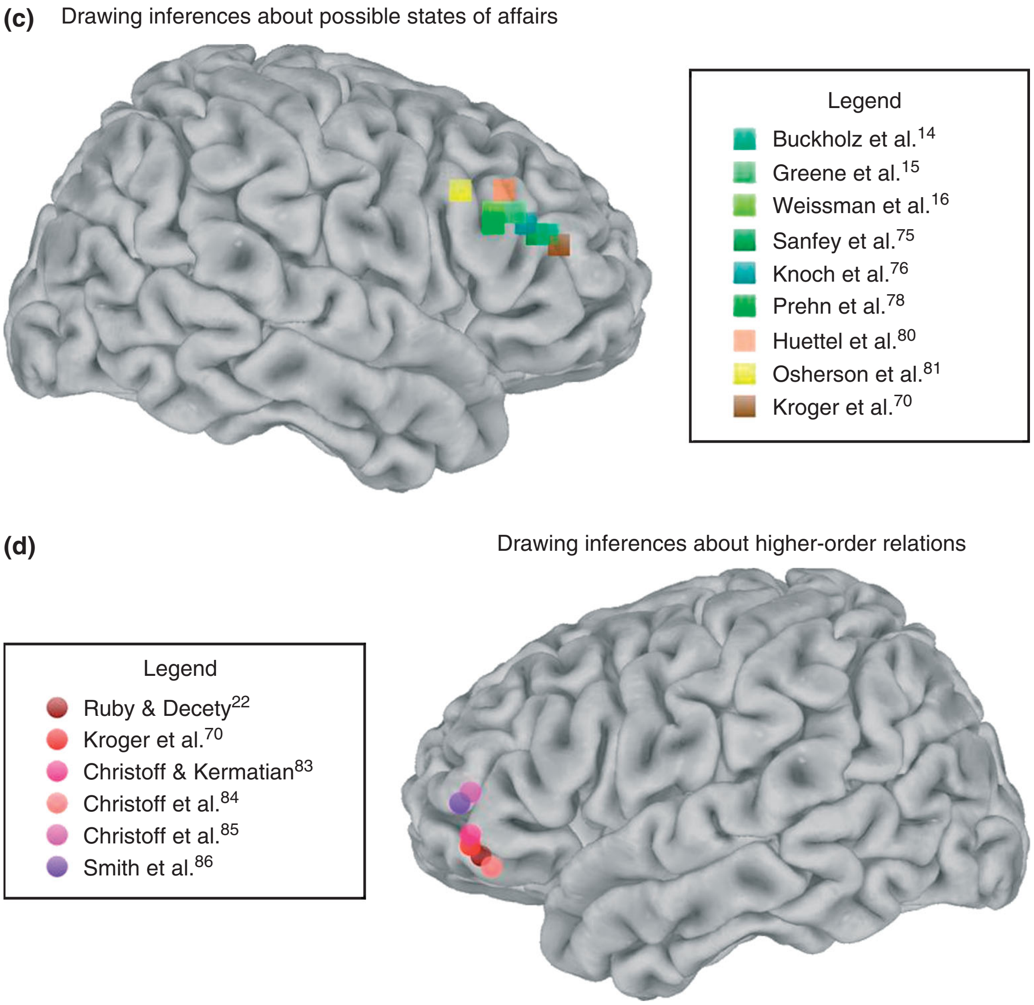


FIGURE 3 | continued.

et al.[76] further demonstrated that deactivating this region with repetitive transcranial magnetic stimulation reduced participants' ability to reject unfair offers in the ultimatum game, suggesting that the dlPFC is central for guiding behavior based on evaluations of fairness and permissibility. Additional evidence is provided by Buckholtz et al.,[14] who observed activity within the right dlPFC (area 46) when participants assigned responsibility for crimes and made judgments about appropriate (e.g., equitable or fair) forms of punishment in a legal decision-making task. The work of Greene et al.[15] further suggests that this region is critical for normative evaluations involving conflicting moral goals. These authors employed moral scenarios similar to the famous trolley problem[77] and assessed trials in which participants acted in the interest of greater aggregate welfare at the expense of personal moral standards. This contrast revealed reliable activation within the right dlPFC (area 46), suggesting that this region is critical for evaluating the permissibility or fairness of behaviors that conflict with personal moral standards (for additional evidence, see Refs 16,78).

Further evidence to support this framework derives from the decision neuroscience literature, which demonstrates the involvement of the dlPFC when drawing conclusions about possible or permissible states of affairs. In contrast to deductive inference, conclusions about possible courses of action reflect uncertainty concerning the actions that 'should' be taken and/or the consequences that 'might' follow, and are referred to as *inductive inferences*. Volz et al.[79] found that activation within the right dlPFC (area 9) increased parametrically with the degree of uncertainty held by the participant (see also Ref 80). Furthermore, Osherson et al.[81] observed preferential recruitment of the right dlPFC (area 46) when performance on an inductive reasoning task was directly compared with a matched deductive inference task, supporting the role of this region for reasoning about possible (rather than necessary) states of affairs.

Anterolateral Prefrontal Cortex

A large body of social neuroscience evidence demonstrates that the alPFC (areas 10 and 11)—and the orbitofrontal cortex (OFC) more broadly—is central for social cognition (Figure 3d). Studies of patients with lesions confined to the OFC have reported impairments in a wide range of social functions, including the regulation and control of social responses, the perception and integration of social cues, and perspective taking.[19–22] Recent evidence from Stone et al.[35] further demonstrates that patients with orbitofrontal damage produced selective impairments in reasoning about social contracts, supporting the proposed role of the PFC in social exchange. Bechara et al.[20] observed profound deficits in the ability of orbitofrontal patients to represent and integrate social and emotional knowledge in the service of decision making. Converging evidence is provided by LoPresti et al.,[21] who demonstrated that the left alPFC (area 11) mediates the integration of multiple social cues (i.e., emotional expression and personal identity), further suggesting that this region supports the integration of multiple classes of social knowledge. Further functional neuroimaging (fMRI) evidence was provided by Moll et al.,[82] who reported bilateral recruitment of the OFC (area 11) during a social decision-making task when participants had to evaluate the social contributions of a charitable organization.

Additional support derives from the decision neuroscience literature, which demonstrates that progressively anterior subregions of the lateral PFC (areas 10 and 11) are associated with higher order processing requirements for thought and action.[65–67] Ramnani and Owen[53] reviewed contemporary research and theory investigating the cognitive functions of the alPFC, concluding that this region is central for integrating the outcomes of multiple cognitive operations, consistent with the predicted role of the alPFC for representing higher order inferences that incorporate both necessary and possible states of affairs (for representative findings, see Refs 70,83–86).

Summary

We have reviewed converging lines of evidence to support an evolutionarily adaptive neural architecture for goal-directed social behavior within the lateral PFC, drawing upon recent theoretical developments in evolutionary psychology and emerging neuroscience evidence investigating the biology, evolution, ontogeny, and cognitive functions of this region. We have surveyed a broad range of social and decision neuroscience evidence demonstrating that the lateral PFC mediates behavior-guiding principles for specific classes of inference, with the vlPFC recruited when drawing inferences about necessary (obligatory or prohibited) courses of action, engagement of the dlPFC when reasoning about possible (permissible) behavior, and the alPFC recruited when both categories of inference are utilized (Figure 3a).

FROM BEHAVIOR-GUIDING PRINCIPLES OF HUMAN INFERENCE TO MORAL BELIEF SYSTEMS

We propose that the inferential architecture of the lateral PFC provides the basis for moral systems of value and belief and therefore embodies two major categories of inference. According to this framework, moral beliefs represent (1) behavior-guiding principles for necessary courses of action (obligation or prohibition) that convey an almost immediate sense of right or wrong (i.e., moral appraisal). In addition, moral beliefs depend upon (2) behavior-guiding principles for possible (permissible) courses of action that enable deliberative reasoning about the fairness of observed behavior (i.e., deliberative moral reasoning). Finally, incorporating both moral appraisal and deliberation will recruit behavior-guiding principles representing both necessary and possible courses of action. Thus, we predict that moral judgment will preferentially recruit functionally specialized regions of the lateral PFC, with the involvement of the vlPFC during moral appraisal, the recruitment of the dlPFC for deliberative moral reasoning, and activation in the alPFC for higher order inferences that incorporate both components of moral judgment.

One prominent view of the mechanisms underlying human moral beliefs is compatible with our proposal, claiming that moral judgment depends on two systems of thought, an intuitive system and a deliberative system.[87–89] Dual process theories of moral judgment are consistent with the extensive application of dual system frameworks in the fields of human inference, judgment, and decision making.[90–93] In the context of moral judgment, the intuitive system generates an immediate sense of right or wrong, and the deliberative system supports reasoning about the permissibility or fairness of observed behavior.

Dual process theories of moral judgment further propose that an initial moral appraisal biases input into the deliberative process.[94] According to the simulation framework, our initial moral appraisals (e.g., 'thou shall not kill') are represented within the vlPFC and inform deliberative components of moral judgment within the dlPFC (e.g., regarding the permissibility or fairness of observed behavior).

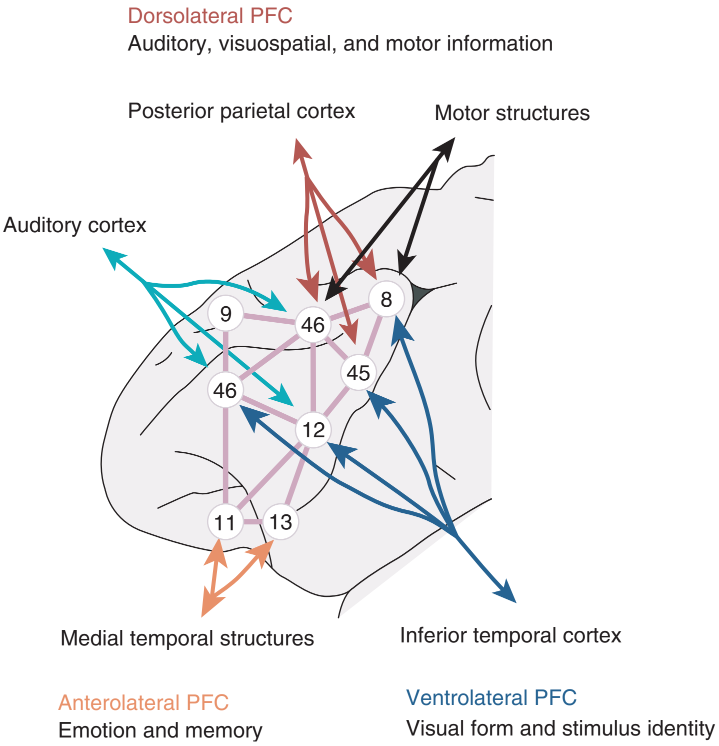


FIGURE 4 | Integrative anatomy of the macaque monkey prefrontal cortex. Numbers refer to subregions within the lateral prefrontal cortex defined by Brodmann. Modified with permission from Miller.[52]

The prominent role of emotional responses in moral appraisals furthermore recruits the alPFC, which is interconnected with the limbic system (via the vmPFC) and additionally supports the integration of intuitive and deliberative components of moral judgment (see Figure 4). Our framework therefore predicts that differences in the relative recruitment of the vlPFC, dlPFC, and alPFC will reflect the respective involvement of moral appraisal, deliberation, or both components in moral judgment. The neuroscience evidence summarized in Figure 3 provides initial support for our proposal, demonstrating that the lateral PFC is functionally specialized to represent these components of moral judgment. The simulation architecture underlying moral beliefs further predicts the recruitment of broadly distributed neural systems, incorporating medial prefrontal[2–4,9]) and posterior knowledge networks (e.g., the right temporal–parietal junction) representing modality-specific components of experience.

CONCLUSION

The reviewed findings elucidate the involvement of the lateral PFC in normative dimensions of social interactions and suggest that simulations provide the basis for moral judgment. In addition, our findings raise further questions for future and emerging programs of neuroscience research. One challenge that awaits future research is to address how behavior-guiding principles for necessary (obligatory and prohibited) and possible (permissible) behavior are represented within dual process theories that distinguish between automatic versus controlled cognitive processes.[5,90] Future research should further investigate the cognitive operations that are performed within the lateral PFC to support human inference. Does this region (1) contain mechanisms that control the recruitment of representations stored in posterior cortices,[52] (2) serve as an integrative hub for synthesizing modality-specific representations,[95] or (iii) store unique forms of knowledge?[10] Future research should also address the biological, developmental, and evolutionary principles that account for the observed lateralization of behavior-guiding principles for necessary (left hemispheric) versus possible (right hemispheric) courses of action (Figure 3). The proposed evolutionary origins and biological basis of behavior-guiding principles for thought and action motivate the question of whether normative standards for human rationality should be constructed from formal mathematical and logical systems, or instead assessed in terms of the evolutionary conditions and ecological contexts that have shaped the development of the human mind.[27–40] Finally, future research should investigate the role of the lateral PFC in the formation of human belief systems, which structure and organize our understanding of the social world. From evolutionarily adaptive social norms represented within the lateral PFC, belief systems for moral,[96,97] ethical, and political[98] thought are constructed. By investigating the origins of this knowledge—assessing the formation of normative principles for goal-directed behavior, and their expression in moral, ethical, and political thought—the burgeoning field of social cognitive neuroscience will continue to advance our understanding of the remarkable cognitive and neural architecture from which uniquely human systems of value and belief emerge.

ACKNOWLEDGEMENTS

The authors are supported by the Intramural Research Program of the National Institute of Neurological Disorders and Stroke. Preparation of this manuscript was based on and adapted from research investigating the role of the prefrontal cortex in goal-directed social behavior.[1–4]

REFERENCES

1. Barbey AK, Grafman J. The prefrontal cortex and goal-directed social behavior. In: Decety J, Cacioppo J, eds. *The Handbook of Social Neuroscience*: Oxford University Press; (in press).
2. Barbey AK, Krueger F, Grafman J. An evolutionarily adaptive neural architecture for social reasoning. *Trends in Neurosciences* 2009, 32:603–610.
3. Barbey AK, Krueger F, Grafman J. Structured event complexes enable predicting the future and explaining the past. In: Bar M, ed. *Predictions in the Brain: Using our Past to Prepare for the Future*: Oxford University Press; (in press).
4. Barbey AK, Krueger F, Grafman J. Structured event complexes in the prefrontal cortex support counterfactual representations for future planning. *Phil Trans R Soc London: Biol Sci* 2009, 364:1291–1300.
5. Lieberman MD. Social cognitive neuroscience: a review of core processes. *Annu Rev Psychol* 2007, 58:259–289.
6. Blakemore SJ, Winston J, Frith U. Social cognitive neuroscience: where are we heading? *Trends Cogn Sci* 2004, 8:215–222.
7. Ochsner KN. Current directions in social cognitive neuroscience. *Curr Opin Neurobiol* 2004, 14:254–258.
8. Amodio DM, Frith CD. Meeting of minds: the medial frontal cortex and social cognition. *Nat Rev Neurosci* 2006, 7:268–277.
9. Krueger F, Barbey AK, Grafman J. The medial prefrontal cortex mediates social event knowledge. *Trends Cogn Sci* 2009, 13:103–109.
10. Wood J, Grafman J. Human prefrontal cortex: processing and representational perspective. *Nat Rev Neurosci* 2003, 4:139–147.
11. Fiddick L, Spampinato MV, Grafman J. Social contracts and precautions activate different neurological systems: an fMRI investigation of deontic reasoning. *NeuroImage* 2005, 28:778–786.
12. Berthoz S, Armony JL, Blair RJR, Dolan RJ. An fMRI study of intentional and unintentional (embarrassing) violations of social norms. *Brain* 2002, 125:1696–1708.
13. Rilling JK, Goldsmith DR, Glenn AL, Jairam MR, Elfenbein HA, et al. The neural correlates of the affective response to unreciprocated cooperation. *Neuropsychologia* 2008, 465:1256–1266.
14. Buckholtz JW, Asplund CL, Dux PE, Zald DH, Gore JC, et al. The neural correlates of third-party punishment. *Neuron* 2008, 60:930–940.
15. Greene JD, Nystrom LE, Engell AD, Darley JM, Cohen JD. The neural bases of cognitive conflict and control in moral judgment. *Neuron* 2004, 44:389–400.
16. Weissman DH, Perkins AS, Woldorff MG. Cognitive control in social situations: a role for the dorsolateral prefrontal cortex. *NeuroImage* 2008, 40:955–962.
17. Damasio AR, Tranel D, Damasio H. Individuals with sociopathic behavior caused by frontal damage fail to respond autonomically to social stimuli. *Behav Brain Res* 1990, 41:81–94.
18. Bechara A, Damasio AR, Damasio H, Anderson SW. Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition* 1994, 50:7–15.
19. Rolls ET, 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* 1994, 57:1518–1524.
20. Bechara A, Damasio H, Damasio A. Emotion, decision making, and the orbitofrontal cortex. *Cereb Cortex* 2000, 10:1047–3211.
21. LoPresti ML, Schon K, Tricarico MD, Swisher JD, Celone KA, et al. Working memory for social cues recruits orbitofrontal cortex and amygdala: a functional

magnetic resonance imaging study of delayed matching to sample for emotional expressions. *J Neurosci* 2008, 28:3718–3728.

22. Ruby P, Decety J. How would you feel versus how do you think she would feel? A neuroimaging study of perspective-taking with social emotions. *J Neurosci* 2004, 16:988–999.

23. Cohen S. Social relationships and health. *Am Psychol* 2004, 59:676–684.

24. Silk JB, Alberts SC, Altmann J. Social bonds of female baboons enhance infant survival. *Science* 2003, 302:1231–1234.

25. Isaac G. The food-sharing behavior of proto human hominids. *Sci Am* 1978, 238:90–108.

26. Brosnan SF, de Waal FBM. Monkeys reject unequal pay. *Nature* 2003, 425:297–299.

27. Maynard Smith J. *Evolution and the Theory of Games*. Cambridge, England: Cambridge University Press; 1982.

28. Cosmides, L. (1985). *Deduction or Darwinian algorithms? An explanation of the "elusive" content effect on the Wason selection task*. Doctoral dissertation, Department of Psychology, Harvard University.

29. Cosmides L. The logic of social exchange: has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition* 1989, 31:187–276.

30. Tooby J, Cosmides L. Friendship and the banker's paradox: other pathways to the evolution of adaptations for altruism. In: Runciman WG, Maynard Smith J, Dunbar RIM, eds. *Evolution of Social Behaviour Patterns in Primates and Man*, vol. 88. Proceedings of the British Academy; 1996, 119–143.

31. Cosmides L, Tooby J. Evolutionary psychology and the generation of culture: part II. Case study: a computational theory of social exchange. *Ethol Sociobiol* 1989, 10:51–97.

32. Cosmides L, Tooby J. Cognitive adaptations for social exchange. In: Barkow J, Cosmides L, Tooby J, eds. *The Adapted Mind*. New York: Oxford University Press; 1992, 163–228.

33. Cosmides L, Tooby J. Social exchange: The evolutionary design of a neurocognitive system. In: Gazzaniga MS, ed. *The New Cognitive Neurosciences*. 3rd ed. Cambridge, MA: MIT Press; 2005.

34. Fiddick L, Cosmides L, Tooby J. No interpretation without representation: the role of domain-specific representations and inferences in the Wason selection task. *Cognition* 2000, 77:1–79.

35. Stone V, Cosmides L, Tooby J, Kroll N, Knight R. Selective impairment of reasoning about social exchange in a patient with bilateral limbic system damage. *Proc Natl Acad Sci U S A* 2002, 99:11531–11536.

36. Sugiyama L, Tooby J, Cosmides L. Cross-cultural evidence of cognitive adaptations for social exchange among the Shiwiar of Ecuadorian Amazonia. *Proc Natl Acad Sci U S A* 2002, 99:11537–11542.

37. Trivers R. The evolution of reciprocal altruism. *Q Rev Biol* 1971, 46:35–57.

38. Axelrod R, Hamilton WD. The evolution of cooperation. *Science* 1981, 211:1390–1396.

39. Platt R, Griggs R. Darwinian algorithms and the Wason selection task: a factorial analysis of social contract selection task problems. *Cognition* 1993, 48:163–192.

40. Gigerenzer G, Hug K. Domain specific reasoning: social contracts, cheating, and perspective change. *Cognition* 1992, 43:127–171.

41. Wason P. Reasoning. In: Foss BM, ed. *New Horizons in Psychology*. Harmondsworth, England: Penguin; 1966, 135–151.

42. Wason P, Realism and rationality in the selection task. In: Evans JBT, ed. *Thinking and Reasoning: Psychological Approaches*. London: Routledge; 1983, 44–75.

43. Wason P, Johnson-Laird P. *The Psychology of Reasoning: Structure and Content*. Cambridge, MA: Harvard University Press; 1972.

44. Cheng P, Holyoak K. Pragmatic reasoning schemas. *Cognit Psychol* 1985, 17:391–416.

45. Krueger F, Moll J, Zahn R, Heinecke A, Grafman J. Event frequency modulates the processing of daily life activities in human medial prefrontal cortex. *Cereb Cortex* 2007, 17:2346–2353.

46. Krueger F, Spampinato M, Barbey AK, Huey T, Morland T, Grafman J. The role of the frontopolar cortex and inferior parietal cortex in mediating action complexity and duration: A parametric fMRI study. *Neuroreport* 2009, 20:1093–1097.

47. Barsalou LW, Simmons WK, Barbey AK, Wilson CD. Grounding conceptual knowledge in modality-specific systems. *Trends Cogn Sci* 2003, 7:84–91.

48. Zeki S. *A Vision of the Brain*. Cambridge, MA: Blackwell Science; 1993.

49. Barbey AK, Barsalou LW. Reasoning and problem solving: models. In: Squire L, Albright T, Bloom F, Gage F, Spitzer N, eds. *Encyclopedia of Neuroscience*. Oxford: Academic Press; 2009, 35–43.

50. Barsalou LW. Grounded cognition. *Annu Rev Psychol* 2008, 59:617–645.

51. Barsalou LW, Niedenthal PM, Barbey AK, Ruppert J. Social embodiment. In: Ross B, ed. *The Psychology of Learning and Motivation*, vol. 43. San Diego: Academic Press; 2003, 43–91.

52. Miller EK. The prefrontal cortex and cognitive control. *Nat Rev Neurosci* 2000, 1:59–65.

53. Ramnani N, Owen AM. Anterior prefrontal cortex: insights into function from anatomy and neuroimaging. *Nat Rev Neurosci* 2004, 5:184–194.

54. Goldman-Rakic PS. Circuitry of primate prefrontal cortex and regulation of behavior by representational

memory. In: Plum F, ed. *Handbook of Physiology: The Nervous System*. Bethesda, MD: American Physiology Society; 1987, 373–417.

55. Pandya DN, Barnes CL. Architecture and connections of the frontal lobe. In: Perecman E, ed. *The Frontal Lobes Revisited*. New York: The IRBN Press; 1987, 41–72.

56. Fuster JM. *The Prefrontal Cortex*. New York: Raven; 1989.

57. Barbas H, Pandya D. Patterns of connections of the prefrontal cortex in the rhesus monkey associated with cortical architecture. In: Levin HS, Eisenberg HM, Bent AL, eds. *Frontal Lobe Function and Dysfunction*. New York: Oxford University Press; 1991, 35–58.

58. Gil-da-Costa R, Braun A, Lopes M, Hauser MD, Carson RE, et al. Toward an evolutionary perspective on conceptual representation: species-specific calls activate visual and affective processing systems. *Proc Nat Acad Sci U S A* 2004, 101:17516–17521.

59. Barsalou LW. Continuity of the conceptual system across species. *Trends Cogn Sci* 2005, 9:309–311.

60. Anderson JA. *An Introduction to Neural Networks*. Cambridge, MA: MIT Press; 1995.

61. Fuster JM. *The Prefrontal Cortex – Anatomy Physiology, and Neuropsychology of the Frontal Lobe* Philadelphia, PA: Lippincott-Raven; 1997.

62. Flechsig P. Developmental (myelogenetic) localisation of the cerebral cortex in the human subject. *Lancet* 1901, 2:1027–1029.

63. Flechsig P. *Anatomie des Menschlichen Gehirnsund Ruckenmarks auf Myelogenetischer Grundlage*. Leipzig, Thieme, New York: Basic Books; 1920.

64. Santrock JW. *Children*, 8th ed. New York: McGraw-Hill; 2005.

65. Badre D. Cognitive control, hierarchy, and the rostrocaudal organization of the frontal lobes. *Trends Cogn Sci* 2008, 12:193–200.

66. Botvinick MM. Hierarchical models of behavior and prefrontal function. *Trends Cogn Sci* 2008, 12:201–208.

67. Koechlin E, Summerfield C. An information theoretical approach to prefrontal executive function. *Trends Cogn Sci* 2007, 11:229–235.

68. Marsh A, Blair K, Jones M, Soliman N, Blair R. Dominance and submission: the ventrolateral prefrontal cortex and responses to status cues. *J Cogn Neurosci* 2009, 4:713–724.

69. Monti MM, Osherson DN, Martinez MJ, Parsons LM. Functional neuroanatomy of deductive inference: a language-independent distributed network. *NeuroImage* 2007, 37:1005–1016.

70. Kroger JK, Nystrom LE, Cohen JD, Johnson-Laird PN. Distinct neural substrates for deductive and mathematical processing. *Brain Res* 2008, 1243:86–103.

71. Heckers S, Zalesak M, Weiss AP, Ditman T, Titone D. Hippocampal activation during transitive inference in humans. *Hippocampus* 2004, 14:153–162.

72. Goel V, Buchel C, Frith C, Dolan R. Dissociation of mechanisms underlying syllogistic reasoning. *NeuroImage* 2000, 12:504–514.

73. Goel V, Dolan RJ. Differential involvement of left prefrontal cortex in inductive and deductive reasoning. *Cognition* 2004, 93:B109–B121.

74. Noveck IA, Goel V, Smith KW. The neural basis of conditional reasoning with arbitrary content. *Cortex* 2004, 40:613–622.

75. Sanfey AG, Rilling JK, Aronson JA, Nystrom LE, Cohen JD. The neural basis of decision making in the ultimatum game. *Science* 2003, 300:1755–1758.

76. Knoch D, Pascual-Leone A, Meyer K, Treyer V, Fehr E. Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science* 2006, 314:829–832.

77. Thomson JJ. Killing, letting die, and the trolley problem. *Monist* 1976, 59:204–217.

78. Prehn K, Wartenburger I, Meriau K, Scheibe C, Goodenough OR, et al. Individual differences in moral judgment competence influence neural correlates of socio-normative judgments. *Soc Cogn Affect Neurosci* 2008, 3:33–46.

79. Volz KG, Schubotz RI, von Cramon Y. Why am I unsure? Internal and external attributions of uncertainty dissociated by fMRI. *NeuroImage* 2004, 21: 848–857.

80. Huettel SA, Song AW, McCarthy G. Decisions under uncertainty: probabilistic context influences activation of prefrontal and parietal cortices. *Eur J Neurosci* 2005, 25:3304–3311.

81. Osherson DN, Perani D, Cappa S, Schnur T, Grassi F, et al. Distinct brain loci in deductive versus probabilistic reasoning. *Neuropsychologia* 1998, 36:369–376.

82. Moll J, Krueger F, Zahn R, Pardini M, de Oliveira-Souza R, et al. Human fronto-mesolimbic networks guide decisions about charitable donation. *Proc Natl Acad Sci U S A* 2006, 103:15623–15628.

83. Christoff K, Keramatian K. Abstraction of mental representations: theoretical considerations and neuroscientific evidence. In: Bunge SA, Wallis JD, eds. *The Neuroscience of Rule-Guided Behavior*: Oxford University Press; 2007.

84. Christoff K, Prabhakaran V, Dorfman J, Zhao Z, Kroger JK, et al. Rostrolateral prefrontal cortex involvement in relational integration during reasoning. *NeuroImage* 2001, 14:1136–1149.

85. Christoff K, Ream JM, Geddes LPT, Gabrieli JDE. Evaluating self-generated information: anterior prefrontal contributions to human cognition. *Behav Neurosci* 2003, 117:1161–1168.

86. Smith R, Keramatian K, Christoff K. Localizing the rostrolateral prefrontal cortex at the individual level. *NeuroImage* 2007, 36:1387–1396.

87. Cushman F, Young L, Hauser M. The role of conscious reasoning and intuition in moral judgments: testing three principles of harm. *Psychol Sci* 2006, 17:1082–1089.

88. Greene JD. The secret joke of Kant's soul. In: Sinnott-Armstrong W, ed. *Moral Psychology: The Neuroscience of Morality*, vol. 3. Cambridge, MA: MIT Press; 2008.

89. Pizarro D, Bloom P. The intelligence of moral intuitions: comment on Haidt (2001). *Psychol Rev* 2003, 110:193–196.

90. Barbey AK, Sloman SA. Base-rate respect: from ecological rationality to dual processes. *Behav Brain Sci* 2007, 30:241–297.

91. Evans JBT. In two minds: dual-process accounts of reasoning. *Trends Cogn Sci* 2003, 7:454–459.

92. Kahneman D, Frederick S. Representativeness revisited: attribute substitution inintuitive judgment. In: Gilovich T, Griffin DW, Kahneman D, eds. *Heuristics and Biases: The Psychology of Intuitive Judgment*. New York: Cambridge University Press; 2002.

93. Stanovich KE, West RF. Individual differences in reasoning: implications for the rationality debate. *Behav Brain Sci* 2000, 23:645–726.

94. Haidt J. The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychol Rev* 2001, 108:814–834.

95. Pessoa L. On the relationship between emotion and cognition. *Nat Rev Neurosci* 2008, 9:148–158.

96. Moll J, Zahn R, de Oliveira-Souza R, Krueger F, Grafman J. The neural basis of human moral cognition. *Nat Rev Neurosci* 2005, 6:799–809.

97. Kapogiannis D, Barbey AK, Su M, Zamboni G, Krueger F, et al. Cognitive and neural foundations of religious belief. *Proc Nat Acad Sci U S A* 2009, 106:4876–4881.

98. Zamboni G, Gozzi M, Krueger F, Duhamel J, Sirigu A, Grafman J. Individualism, conservatism, and radicalism as criteria for processing political beliefs: a parametric fMRI study. *Soc Neurosci* (in press).

FURTHER READING

Greene JD, Sommerville R, Nystrom L, Darley JM, Cohen JD. An fMRI investigation of emotional engagement in moral judgment. *Science* 2001, 293:2105–2108.

Foot P. *Virtues and Vices and Other Essays in Moral Philosophy*. Berkeley: University of California Press; 1978.

Malle BF. Folk explanations of intentional action. In: Malle BF, Moses LJ, Baldwin DA, eds. *Intentions and Intentionality: Foundations of Social Cognition*. Cambridge, MA: MIT Press; 2001.

Malle BF, Knobe J. The folk concept of intentionality. *J Exp Soc Psychol* 1997, 2:101–121.

Mikhail J. *Rawls' Linguistic Analogy: A Study of the 'Generative Grammar' Model of Moral Theory Described by John Rawls in 'A Theory of Justice.'*, PhD Dissertation, Cornell University, 2000.

Nowak MA. Five rules for the evolution of cooperation. *Science* 2006, 314:1560–1563.

Spranca M, Minsk E, Baron J. Omission and commission in judgment and choice. *J Exp Soc Psychol* 1991, 27:76–105.

Young L, Cushman F, Hauser M, Saxe R. The neural basis of the interaction between theory of mind and moral judgment. *Proc Natl Acad Sci U S A* 2007, 104:8235–8240.