Review

An evolutionarily adaptive neural architecture for social reasoning

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

[1] Cognitive Neuroscience Section, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD, USA
[2] Department of Psychology, Georgetown University, Washington, DC, USA

Recent progress in cognitive neuroscience highlights the involvement of the prefrontal cortex (PFC) in social cognition. Accumulating evidence demonstrates that representations within the lateral PFC enable people to coordinate their thoughts and actions with their intentions to support goal-directed social behavior. Despite the importance of this region in guiding social interactions, remarkably little is known about the functional organization and forms of social inference processed by the lateral PFC. Here, we introduce a cognitive neuroscience framework for understanding the inferential architecture of the lateral PFC, drawing upon recent theoretical developments in evolutionary psychology and emerging neuroscience evidence about how this region can orchestrate behavior on the basis of evolutionarily adaptive social norms for obligatory, prohibited and permissible courses of action.

Architect of the social mind

Evolution has fundamentally shaped the architecture of the mind, producing cognitive and neural mechanisms that are designed to solve adaptive problems encountered by our human ancestors. Throughout evolutionary history, a foremost adaptive challenge for our species was living and interacting with people – learning to select mates, form alliances and compete for limited resources. Our human ancestors also needed to obey social norms and standard of conduct, as violations of these rules might have restricted the formation of organized social groups and might have been severely punished. Accordingly, just as the brain has evolved mechanisms for perception, memory, language and thought, it is probable that there are also evolutionarily adaptive mechanisms that enable humans to coexist with others.

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 [1–3]). This interdisciplinary enterprise has made considerable progress in understanding the involvement of the prefrontal cortex (PFC) in social cognition [4–11]. Accumulating evidence indicates 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 [12–23]. Despite the pivotal role of this region in guiding social interactions, fundamental questions remain concerning the functional organization and forms of social inference processed by the lateral PFC. Here, we develop an integrative cognitive neuroscience framework for understanding the inferential architecture 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 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, ethical, legal and political 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 (for representative findings from behavioral economics, see Refs [24–29]). From our earliest ancestors to present day, social exchange has facilitated access to sustenance, protection and mates, and enabled people to

Corresponding author: Grafman, J. (GrafmanJ@ninds.nih.gov).

0166-2236/$ – see front matter. Published by Elsevier Ltd. doi:10.1016/j.tins.2009.09.001 Available online xxxxxx

live healthier and longer lives [30,31]. 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 [32,33]. Evolutionary psychologists have proposed that social exchange embodies cognitive mechanisms 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 [34–44].

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 [35–39]. 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 [40,41]. Accumulating evidence supports this proposal, demonstrating that the mind embodies functionally specialized cognitive mechanisms for detecting cheaters [35–39] 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 type 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 is prohibited).

A primary method for investigating conditional reasoning about social contracts is the Wason four-card selection task [45]. 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 [35–39,46]. 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) [35–39]. Cognitive experiments have demonstrated that this improved level of performance is sensitively regulated by the series of variables expected if this were a system optimally designed to reason about obligatory and prohibited forms of social behavior, rather than to support a broader class of inferences [35–39,43,46].

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 (i) *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 (ii) *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 [34,36–43]. 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 [34–43], 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.

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.

Simulation theory of prefrontal cortex function

One of the great mysteries of brain function concerns how coordinated, purposeful behavior arises from neural states. How are people able to orchestrate their thoughts and actions in concert with their intentions to support goal-directed social behavior? An emerging body of evidence suggests that this capacity 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 (for empirical reviews from neuroscience and psychology, see Refs [5–11,47–49]). 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. [50]). 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 [5–11,47–49].

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 the lateral PFC comprises neurons that are exquisitely sensitive to behaviorally informative associations (for a review, see Ref. [51]).

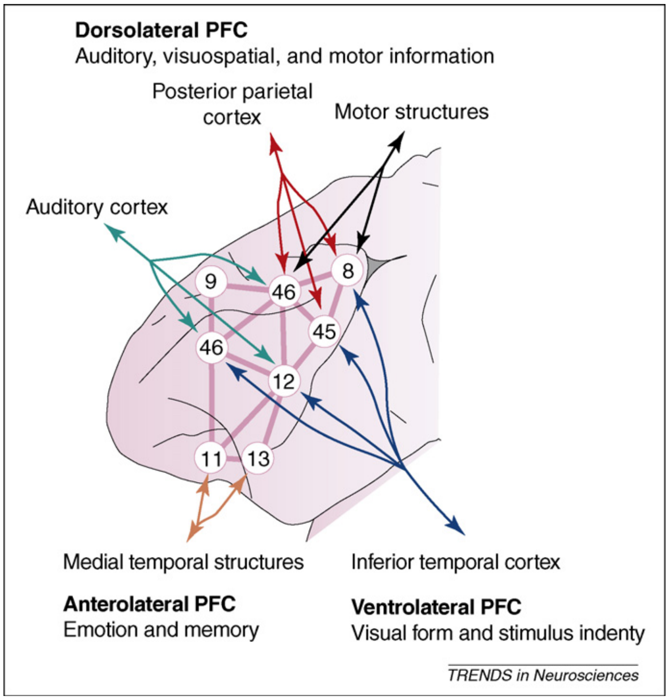


Figure 1. Integrative anatomy of the macaque monkey prefrontal cortex (PFC). Numbers refer to subregions within the lateral PFC defined by Brodmann. Modified with permission from Ref. [51].

This work has focused on the lateral PFC because it represents a site of convergence of the information needed to synthesize multimodal information from a wide range of brain systems. The lateral PFC consists of three major subregions that emphasize processing of particular information based on their interconnections with specific cortical regions (Figure 1).

Ventrolateral PFC (vlPFC) is heavily interconnected with cortical areas 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. Dorsolateral PFC (dlPFC) is interconnected with cortical areas for processing auditory, visuospatial and motor information, enabling the regulation and control of responses to environmental stimuli. Finally, anterolateral PFC (alPFC) is indirectly connected (via the ventromedial PFC) with limbic structures that process internal information, such as emotion, memory and reward [52–55]. The lateral PFC therefore enables the integration and synthesis of information across this broadly distributed network of modal brain regions [7].

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 [5–10,47,48]. For example, on entering a familiar situation and recognizing it, a simulation that represents the situation becomes active. Typically the entire situation is not perceived initially. A relevant person, setting or event might 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 (i) situation (including agents, objects, actions, mental states and background settings) and (ii) causal and associative relations that hold among its elements [10,56–59]. 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 for observed behavior [5–7].

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. We propose that behavior-guiding principles for social inference are mediated by deeply entrenched simulations, for which learned associations are the product of evolutionarily adaptive cognitive and neural mechanisms [34–43].

The observed role of simulation mechanisms for social inference in non-human primates supports this approach [60], suggesting that modality-specific simulations represent continuity of social information processing across the species [61]. 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 [62]. 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 'might' take. The unfolding of inferences about necessary and possible states of affairs – realized as a simulation – represent behavior-guiding principles for the orchestration of social thought and action [5–11,47,48]. The recruitment of specific lateral PFC subregions for social inference is determined by the evolution, development, hierarchical structure and anatomical connectivity of the PFC (Box 1). 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 (i) the involvement of vlPFC when reasoning about necessary (obligatory or prohibited) courses of action; (ii) the recruitment of dlPFC for drawing inferences about possible (permissible) states of affairs; and (iii) activation in alPFC for higher-order inferences that incorporate both categories of knowledge (Figure 2). The simulation architecture underlying these forms of inference further predicts the recruitment of broadly distributed neural systems, incorporating medial prefrontal [4–6,9–11] and posterior knowledge networks representing modality-specific components of experience. Our review examines functional neuroimaging (fMRI) data in healthy volunteers, neuropsychological studies of brain-injured patients and repetitive transcranial

Box 1. An evolutionarily adaptive neural architecture for social reasoning

The functional organization of the lateral PFC is determined by the evolution, development, hierarchical structure and anatomical connectivity of the PFC (Figure I).

Evolution and development

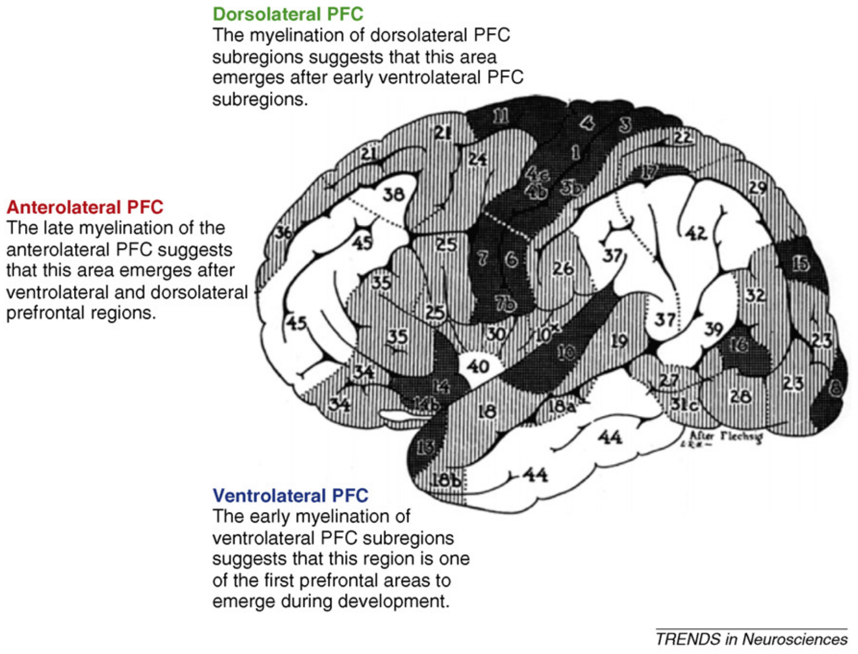
Research investigating the evolution and ontogeny of the PFC demonstrates that the lateral PFC initially emerged from ventrolateral prefrontal regions, followed by dorsolateral and then anterolateral cortices [54,90]. From an evolutionary perspective, the emergence of lateral PFC subregions reflects their relative priority for the formation of organized social groups, with 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 dlPFC. Finally, the evolution of 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 [91].

Hierarchical organization

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 [78–81]). 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.

Anatomical connectivity

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.* [92] 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 [34–43], 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).



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 prefrontal regions.

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

TRENDS in Neurosciences

Figure Box 1. Ontogenetic map of the prefrontal cortex (PFC) according to Flechsig [90]. The numeration of the areas indicates the order of their myelination. Modified with permission from Ref. [90].

magnetic stimulation (rTMS) evidence, identifying the recruitment of distinct and often lateralized subregions of the lateral PFC for alternative forms of human inference.

Ventrolateral prefrontal cortex

An increasing number of fMRI studies have shown that social norms for necessary (obligatory or prohibited) courses of action are represented by vlPFC (areas 44, 45 and 47; Figure 2b). Fiddick *et al.* [12] 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.* [13], who demonstrated the recruitment of left vlPFC (area 47) when participants detected violations of social norms stories representing obligatory and prohibited courses of action (e.g. the decision to “spit out food made by the host”).

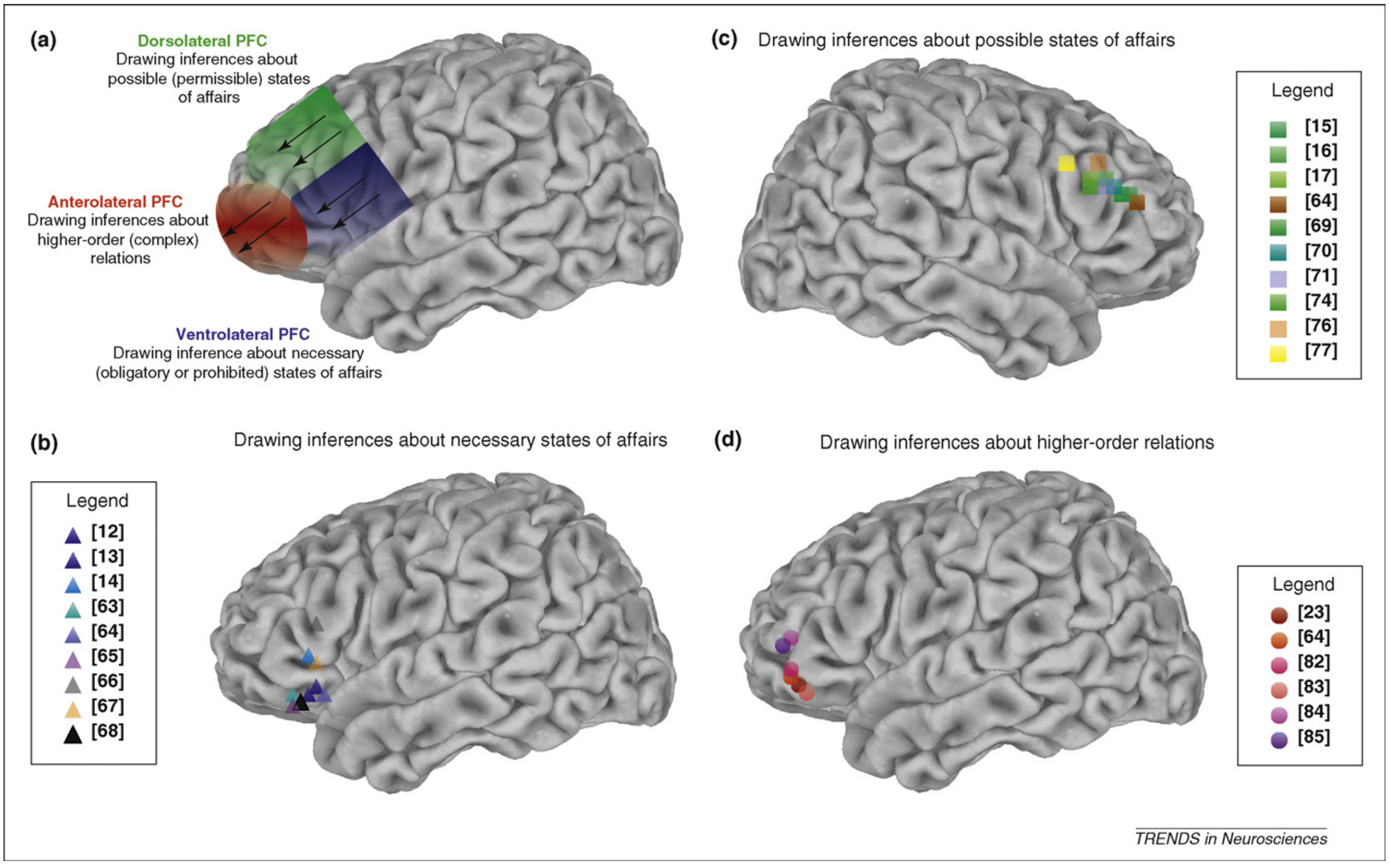

(a)
Dorsolateral PFC
Drawing inferences about possible (permissible) states of affairs
Anterolateral PFC
Drawing inferences about higher-order (complex) relations
Ventrolateral PFC
Drawing inference about necessary (obligatory or prohibited) states of affairs
(c) Drawing inferences about possible states of affairs
Legend
■ [15]
■ [16]
■ [17]
■ [64]
■ [69]
■ [70]
■ [71]
■ [74]
■ [76]
■ [77]
(b) Drawing inferences about necessary states of affairs
Legend
▲ [12]
▲ [13]
▲ [14]
▲ [63]
▲ [64]
▲ [65]
▲ [66]
▲ [67]
▲ [68]
(d) Drawing inferences about higher-order relations
Legend
● [23]
● [64]
● [82]
● [83]
● [84]
● [85]
TRENDS in Neurosciences

Figure 2. An evolutionarily adaptive neural architecture for social reasoning. **(a)** The functional organization of the lateral prefrontal cortex (PFC) which is symmetric in both hemispheres. **(b)**, **(c)**, and **(d)** illustrate supportive evidence. **(b)** and **(d)** illustrate the left hemisphere, **(c)** depicts the right hemisphere.

Similarly, Rilling *et al.* [14] 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 vlPFC when drawing conclusions that necessarily follow from the truth of the premises, that is, 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.* [63] controlled for these sources of variability and provided evidence that 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.* [64] controlled for the complexity and type of calculations that were performed and also observed activation within left vlPFC (areas 44 and 45) for deductive reasoning [65]. Additional supporting data are provided by Goel *et al.* [66,67], who have consistently observed activation within 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.* [68] 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 rule.

Dorsolateral prefrontal cortex

Accumulating evidence demonstrates that dlPFC (areas 46 and 9) represents behavior-guiding principles for evaluating the permissibility or fairness of observed behavior (Figure 2c). An early study by Sanfey *et al.* [69] reported activity within right dlPFC (area 46) when participants evaluated the fairness of an offer in an ultimatum game. Knoch *et al.* [70] further demonstrated that deactivating this region with rTMS reduced participants ability to reject unfair offers in the ultimatum game, suggesting that dlPFC is central for guiding behavior based on evaluations of fairness and permissibility. Additional evidence is provided by Buckholtz *et al.* [15], who observed activity within 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. Similarly, Spitzer *et al.* [71] demonstrated that social norm compliance recruits a lateral prefrontal network, with right dlPFC (area 46) mediating subjective representations of punishment enforced for violating socially appropriate forms of behavior (for additional discussion, see Ref. [72]). The work of Greene *et al.* [16] further suggests that this region is involved in normative evaluations involving conflicting moral goals. These authors employed moral scenarios similar to the famous trolley problem [73] 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 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 [17,74]).

Further evidence to support this framework derives from the decision neuroscience literature, which demonstrates the involvement of 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.* [75] found that activation within right dlPFC (area 9) increased parametrically with the degree of uncertainty held by the participant [76]. Furthermore, Osherson *et al.* [77] observed preferential recruitment of right dlPFC (area 46) when performance on an inductive reasoning task was directly compared to 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 neuroscience evidence demonstrates that alPFC (areas 10 and 11) – and the orbitofrontal cortex (OFC) more broadly – is central for social cognition (Figure 2d). 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 [20–23]. Evidence from Stone *et al.* [39] further demonstrates that patients with orbitofrontal damage produced selective impairments in reasoning about social contracts, supporting the proposed role of the PFC in goal-directed social behavior. Bechara *et al.* [21] observed profound deficits in the ability of OFC patients to represent and integrate social and emotional knowledge in the service of decision-making. Converging evidence is provided by LoPresti *et al.* [22], who demonstrated that 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.

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 [78–80]. Ramnani and Owen [81] reviewed contemporary research and theory investigating the cognitive functions of alPFC, concluding that this region is central for integrating the outcomes of multiple cognitive operations, consistent with the predicted role of alPFC for representing higher-order inferences that incorporate both necessary and possible states of affairs (for representative findings, see Refs [64,82–85]).

Toward an integrative theory of human inference, value and belief

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

Box 2. Outstanding questions

- What role does the lateral PFC play in the formation of human belief systems? In what manner does the neural and computational architecture of this region constrain the form and expression of moral [87–89], ethical and political beliefs? How do these constraints in turn shape the formation of society and invariant properties of human culture?
- What cognitive operations are implemented within the lateral PFC to support human inference? Does this region (i) serve as an integrative hub for the synthesis of modality-specific representations [93]; (ii) contain mechanisms that control the selection and processing of modality-specific knowledge [51]; or (iii) incorporate both representational and processing functions [5–10,47,48].
- In what sense and to what degree does the inferential architecture of the lateral PFC represent intuitive, affective and deliberative processes? How do these decision components together contribute to human reasoning?
- How are inferences concerning necessary and possible states of the world represented within dual process theories that distinguish between the neural and computational architecture of intuitive versus deliberative processes [1,57]?
- The involvement of the lateral PFC for processing social and non-social knowledge raises the question of whether behavior-guiding principles for social inference are mediated by domain-specific (rather than domain-general) neural systems (for recent neuroscience reviews, see Refs [9,10]).
- What evolutionary and biological principles account for the observed lateralization of inferences concerning necessary (left hemispheric) versus possible (right hemispheric) courses of action (Figure 2)?
- The reviewed evidence derives primarily from functional neuroimaging studies (fMRI) that establish an association between specific lateral PFC subregions and forms of reasoning. The necessity of this region and its precise contribution to the neural systems mediating human inference, however, remains controversial [94], raising the question of how the involvement of lateral PFC should be characterized: Is this region associated with and/or necessary for the reviewed categories of inference?
- Evolutionary psychologists have identified a broad range of evolutionarily adaptive cognitive heuristics that support decision-making (for a review, see Ref. [95]), motivating further research and theory investigating how cognitive heuristics are implemented within the neural and computational architecture of the lateral PFC.
- The evolutionary origins and biological bases of behavior-guiding principles for human inference 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 inferential architecture of the human mind [34–43].

are utilized (Figure 2). These findings set the stage for new approaches to understanding human thought – in terms of its evolutionarily adaptive functional organization and the constraints it places on the formation of human systems of value and belief – from the level of genes and neurons to thought and behavior.

Nature provides a seemingly unbounded variety of complex functional structures, from mechanisms for social behavior in insects to the neural architecture that enables social reasoning in humans. Until recently, variation was believed to emerge from random processes, with adaptations shaped solely by natural selection. New molecular approaches have sharpened our understanding of the sources of variation and how developmental programs constrain evolutionary processes, yielding a restricted range of possible adaptations [86]. Our review elucidates the neural architecture that has emerged from these processes to support social reasoning, identifying functionally organized machinery that is too improbably well ordered to have arisen by chance (Figure 2).

An evolutionarily adaptive neural architecture for human inference raises new possibilities for understanding questions of evolution and human intelligence (Box 2), suggesting that research and theory from anthropology, evolutionary psychology, social psychology, political science, behavioral economics and decision neuroscience can be productively combined to understand one of the most profound problems of intellectual life: how humans evolved the capacity to represent and reason about necessity and possibility, and to productively apply this knowledge to construct systems of moral [87–89], ethical, legal and political belief. By investigating the evolutionary origins of this knowledge – assessing the formation of normative principles for social behavior and their extraordinary range of cultural expression – 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.

Acknowledgement
The authors are supported by the Intramural Research Program of the National Institute of Neurological Disorders and Stroke. We are grateful to Lawrence W. Barsalou, Denise Cummins, Amita Srivastava and three anonymous reviewers for helpful comments on earlier drafts.

References

1 Lieberman, M.D. (2007) Social cognitive neuroscience: a review of core processes. *Annu. Rev. Psychol.* 58, 259–289

2 Blakemore, S.J. *et al.* (2004) Social cognitive neuroscience: where are we heading? *Trends Cogn. Sci.* 8, 216–222

3 Ochsner, K.N. (2004) Current directions in social cognitive neuroscience. *Curr. Opin. Neurobiol.* 14, 254–258

4 Amodio, D.M. and Frith, C.D. (2006) Meeting of minds: the medial frontal cortex and social cognition. *Nat. Rev. Neurosci.* 7, 268–277

5 Barbey, A.K. *et al.* (2009) Structured event complexes in the medial prefrontal cortex support counterfactual representations for future planning. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 1291–1300

6 Barbey, A.K. *et al.* (2009) Structured event complexes and mental models for counterfactual inference. In *Predictions in the Brain: Using our Past to Prepare for the Future* (Bar, M., ed.), Oxford University Press

7 Barbey, A.K. and Grafman, J. An integrative cognitive neuroscience framework for social reasoning and moral judgment. *Wiley Interdiscip. Rev. Cogn. Sci*. (in press)

8 Barbey, A.K. and Grafman, J. The prefrontal cortex and goal-directed social behavior. In *The Handbook of Social Neuroscience* (Decety, J. and Cacioppo, J., eds), Oxford University Press (in press)

9 Barbey, A.K. and Barsalou, L.W. (2009) Reasoning and problem solving: models. In *Encyclopedia of Neuroscience* (Squire, L. *et al.*, eds), pp. 35–43, Academic Press

10 Patterson, R. and Barbey, A.K. Causal simulation theory: an integrative cognitive neuroscience framework for causal reasoning. In *Neural Basis of Belief Systems* (Grafman, J. and Krueger, F., eds), Psychology Press (in press)

11 Krueger, F. *et al.* (2009) The medial prefrontal cortex mediates social event knowledge. *Trends Cogn. Sci.* 13, 103–109

12 Fiddick, L. *et al.* (2005) Social contracts and precautions activate different neurological systems: an fMRI investigation of deontic reasoning. *Neuroimage* 28, 778–786

13 Berthoz, S. *et al.* (2002) An fMRI study of intentional and unintentional (embarrassing) violations of social norms. *Brain* 125, 1696–1708

14 Rilling, J.K. *et al.* (2008) The neural correlates of the affective response to unreciprocated cooperation. *Neuropsychologia* 46, 1256–1266

15 Buckholtz, J.W. *et al.* (2008) The neural correlates of third-party punishment. *Neuron* 60, 930–940

16 Greene, J.D. *et al.* (2004) The neural bases of cognitive conflict and control in moral judgment. *Neuron* 44, 389–400

17 Weissman, D.H. *et al.* (2008) Cognitive control in social situations: a role for the dorsolateral prefrontal cortex. *Neuroimage* 40, 955–962

18 Damasio, A.R. *et al.* (1990) Individuals with sociopathic behavior caused by frontal damage fail to respond autonomically to social stimuli. *Behav. Brain Res.* 41, 81–94

19 Bechara, A. *et al.* (1994) Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition* 50, 7–15

20 Rolls, E.T. *et al.* (1994) Emotion-related learning in patients with social and emotional changes associated with frontal lobe damage. *J. Neurol. Neurosurg. Psychiatry* 57, 1518–1524

21 Bechara, A. *et al.* (2000) Emotion, decision making and the orbitofrontal cortex. *Cereb. Cortex* 10, 295–307

22 LoPresti, M.L. *et al.* (2008) 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.* 28, 3718–3728

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

24 Fehr, E. and Fischbacher, U. (2003) The nature of human altruism. *Nature* 425, 785–791

25 Fehr, E. and Fischbacher, U. (2004) Social norms and human cooperation. *Trends Cogn. Sci.* 8, 185–190

26 Boyd, R. *et al.* (2003) The evolution of altruistic punishment. *Proc. Natl. Acad. Sci. U. S. A.* 100, 3531–3535

27 Panchanathan, K. and Boyd, R. (2004) Indirect reciprocity can stabilize cooperation without the second-order free rider problem. *Nature* 432, 499–502

28 Bowles, S. (2006) Group competition, reproductive leveling, and the evolution of human altruism. *Science* 314, 1569–1572

29 Bernhard, H. *et al.* (2006) Parochial altruism in humans. *Nature* 442, 912–915

30 Cohen, S. (2004) Social relationships and health. *Am. Psychol.* 59, 676–684

31 Silk, J.B. *et al.* (2003) Social bonds of female baboons enhance infant survival. *Science* 302, 1231–1234

32 Isaac, G. (1978) The food-sharing behavior of protohuman hominids. *Sci. Am.* 238, 90–108

33 Brosnan, S.F. and De Waal, F.B. (2003) Monkeys reject unequal pay. *Nature* 425, 297–299

34 Maynard Smith, J. (1982) *Evolution and the Theory of Games*, Cambridge University Press

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

36 Cosmides, L. and Tooby, J. (1992) Cognitive adaptations for social exchange. In *The Adapted Mind* (Barkow, J. *et al.*, eds), pp. 163–228, Oxford University Press

37 Cosmides, L. and Tooby, J. (2005) Social exchange: The evolutionary design of a neurocognitive system. In *The New Cognitive Neurosciences*

(III) (Gazzaniga, M., ed.), In pp. 1295–1308, Cambridge, MA, MIT press
38 Fiddick, L. *et al.* (2000) No interpretation without representation: the role of domain-specific representations and inferences in the Wason selection task. *Cognition* 77, 1–79
39 Stone, V.E. *et al.* (2002) Selective impairment of reasoning about social exchange in a patient with bilateral limbic system damage. *Proc. Natl. Acad. Sci. U. S. A.* 99, 11531–11536
40 Trivers, R. (1971) The evolution of reciprocal altruism. *Q. Rev. Biol.* 46, 35–57
41 Axelrod, R. and Hamilton, W.D. (1981) The evolution of cooperation. *Science* 211, 1390–1396
42 Platt, R.D. and Griggs, R.A. (1993) Darwinian algorithms and the Wason selection task: a factorial analysis of social contract selection task problems. *Cognition* 48, 163–192
43 Gigerenzer, G. and Hug, K. (1992) Domain-specific reasoning: social contracts, cheating, and perspective change. *Cognition* 43, 127–171
44 Pinker, S. (1994) *The Language Instinct*, Harcourt
45 Wason, P.C. (1983) Realism and rationality in the selection task. In *Thinking and Reasoning: Psychological Approaches* (Evans, J.St.B.T., ed.), pp. 44–75, Routledge
46 Cheng, P.W. and Holyoak, K.J. (1985) Pragmatic reasoning schemas. *Cogn. Psychol.* 17, 391–416
47 Barsalou, L.W. *et al.* (2003) Grounding conceptual knowledge in modality-specific systems. *Trends Cogn. Sci.* 7, 84–91
48 Barsalou, L.W. *et al.* (2003) Social embodiment. In *The Psychology of Learning and Motivation* (Ross, B., ed.), pp. 43–91, Academic Press
49 Damasio, A.R. (1989) Time-locked multiregional retroactivation: a systems-level proposal for the neural substrates of recall and recognition. *Cognition* 33, 25–62
50 Zeki, S. (1993) *A Vision of the Brain*, Blackwell Scientific Publications Inc
51 Miller, E.K. (2000) The prefrontal cortex and cognitive control. *Nat. Rev. Neurosci.* 1, 59–65
52 Goldman-Rakic, P.S. (1987) Circuitry of the frontal association cortex and its relevance to dementia. *Arch. Gerontol. Geriatr.* 6, 299–309
53 Pandya, D.N. and Barnes, C.L. (1987) Architecture and connections of the frontal lobe. In *The Frontal Lobes Revisited* (Perecman, E., ed.), pp. 41–72, The IRBN Press
54 Fuster, J.M. (1997) *The Prefrontal Cortex: Anatomy, Physiology, and Neuropsychology of the Frontal Lobe*, Lippincott-Raven
55 Barbas, H. and Pandya, D.N. (1991) Patterns of connections of the prefrontal cortex in the rhesus monkey associated with cortical architecture. In *Frontal Lobe Function and Dysfunction* (Levin, H.S. *et al.*, eds), pp. 35–58, Oxford, Oxford University Press
56 Barbey, A.K. and Wolff, P. (2007) Learning causal structure from reasoning. In *Proceedings of the Twenty-Ninth Annual Conference of the Cognitive Science Society* (McNamara, D. *et al.*, eds), pp. 713–718, Nashville, TN, Cognitive Neuroscience Society
57 Barbey, A.K. and Sloman, S.A. (2007) Base-rate respect: from ecological rationality to dual processes. *Behav. Brain Sci.* 30, 241–254 discussion 255–297
58 Sloman, S.A. *et al.* (2009) A causal model theory of the meaning of “cause”, “enable”, and “prevent”. *Cogn. Sci.* 33, 21–50
59 Chaigneau, S. and Barbey, A.K. (2008) Assessing psychological theories of causal meaning and inference. In *Proceedings of the Thirtieth Annual Conference of the Cognitive Science Society* (Love, B., ed.), pp. 111–116, Austin, TX, Cognitive Science Society
60 Gil-da-Costa, R. *et al.* (2004) Toward an evolutionary perspective on conceptual representation: species-specific calls activate visual and affective processing systems in the macaque. *Proc. Natl. Acad. Sci. U. S. A.* 101, 17516–17521
61 Barsalou, L.W. (2005) Continuity of the conceptual system across species. *Trends Cogn. Sci.* 9, 309–311
62 Anderson, J.A. (1995) *An Introduction to Neural Networks*, MIT Press
63 Monti, M.M. *et al.* (2007) Functional neuroanatomy of deductive inference: a language-independent distributed network. *Neuroimage* 37, 1005–1016
64 Kroger, J.K. *et al.* (2008) Distinct neural substrates for deductive and mathematical processing. *Brain Res.* 1243, 86–103
65 Heckers, S. *et al.* (2004) Hippocampal activation during transitive inference in humans. *Hippocampus* 14, 153–162
66 Goel, V. *et al.* (2000) Dissociation of mechanisms underlying syllogistic reasoning. *Neuroimage* 12, 504–514
67 Goel, V. and Dolan, R.J. (2004) Differential involvement of left prefrontal cortex in inductive and deductive reasoning. *Cognition* 93, B109–B121
68 Noveck, I.A. *et al.* (2004) The neural basis of conditional reasoning with arbitrary content. *Cortex* 40, 613–622
69 Sanfey, A.G. *et al.* (2003) The neural basis of economic decision-making in the Ultimatum Game. *Science* 300, 1755–1758
70 Knoch, D. *et al.* (2006) Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science* 314, 829–832
71 Spitzer, M. *et al.* (2007) The neural signature of social norm compliance. *Neuron* 56, 185–196
72 Haushofer, J. and Fehr, E. (2008) You shouldn’t have: your brain on others’ crimes. *Neuron* 60, 738–740
73 Thomson, J.J. (1976) Killing, letting die, and the trolley problem. *Monist* 59, 204–217
74 Prehn, K. *et al.* (2008) Individual differences in moral judgment competence influence neural correlates of socio-normative judgments. *Soc. Cogn. Affect. Neurosci.* 3, 33–46
75 Volz, K.G. *et al.* (2004) Why am I unsure? Internal and external attributions of uncertainty dissociated by fMRI. *Neuroimage* 21, 848–857
76 Huettel, S.A. *et al.* (2005) Decisions under uncertainty: probabilistic context influences activation of prefrontal and parietal cortices. *J. Neurosci.* 25, 3304–3311
77 Osherson, D. *et al.* (1998) Distinct brain loci in deductive versus probabilistic reasoning. *Neuropsychologia* 36, 369–376
78 Badre, D. (2008) Cognitive control, hierarchy, and the rostro-caudal organization of the frontal lobes. *Trends Cogn. Sci.* 12, 193–200
79 Botvinick, M.M. (2008) Hierarchical models of behavior and prefrontal function. *Trends Cogn. Sci.* 12, 201–208
80 Koechlin, E. and Summerfield, C. (2007) An information theoretical approach to prefrontal executive function. *Trends Cogn. Sci.* 11, 229–235
81 Ramnani, N. and Owen, A.M. (2004) Anterior prefrontal cortex: insights into function from anatomy and neuroimaging. *Nat. Rev. Neurosci.* 5, 184–194
82 Christoff, K. and Keramatian, K. (2007) Abstraction of mental representations: theoretical considerations and neuroscientific evidence. In *The Neuroscience of Rule-Guided Behavior* (Bunge, S.A. and Wallis, J.D., eds), Oxford University Press
83 Christoff, K. *et al.* (2001) Rostrolateral prefrontal cortex involvement in relational integration during reasoning. *Neuroimage* 14, 1136–1149
84 Christoff, K. *et al.* (2003) Evaluating self-generated information: anterior prefrontal contributions to human cognition. *Behav. Neurosci.* 117, 1161–1168
85 Smith, R. *et al.* (2007) Localizing the rostrolateral prefrontal cortex at the individual level. *Neuroimage* 36, 1387–1396
86 Hauser, M.D. (2009) The possibility of impossible cultures. *Nature* 460, 190–196
87 Barsalou, L.W. *et al.* (2005) Embodiment in religious knowledge. *J. Cogn. Cult.* 5, 14–57
88 Moll, J. *et al.* (2005) Opinion: the neural basis of human moral cognition. *Nat. Rev. Neurosci.* 6, 799–809
89 Kapogiannis, D. *et al.* (2009) Cognitive and neural foundations of religious belief. *Proc. Natl. Acad. Sci. U. S. A.* 106, 4876–4881
90 Flechsig, P. (1901) Developmental (myelogenetic) localisation of the cerebral cortex in the human subject. *Lancet* 2, 1027–1029
91 Santrock, J.W. (2005) *Children*, (8th edn), McGraw-Hill Press
92 Marsh, A.A. *et al.* (2009) Dominance and submission: the ventrolateral prefrontal cortex and responses to status cues. *J. Cogn. Neurosci.* 21, 713–724
93 Pessoa, L. (2008) On the relationship between emotion and cognition. *Nat. Rev. Neurosci.* 9, 148–158
94 Volle, E., Kinkingnehun, S., Pochon, J., Mondon, K., de Schotten, M.T., Seassau, M., Duffau, H., Samson, Y., Dubois, B. and Levy, R. (2008) The functional architecture of the left posterior and lateral prefrontal cortex in humans. *Cerebral Cortex* 10, 1093–1103
95 Gigerenzer, G. *et al.* (1999) *Simple Heuristics that Make us Smart*, Oxford University Press