Available online at www.sciencedirect.com
ScienceDirect
Current Opinion in Neurobiology

The functional brain architecture of human morality

Chadd M Funk and Michael S Gazzaniga

Human morality provides the foundation for many of the pillars of society, informing political legislation and guiding legal decisions while also governing everyday social interactions. In the past decade, researchers in the field of cognitive neuroscience have made tremendous progress in the effort to understand the neural basis of human morality. The emerging insights from this research point toward a model in which automatic processing in parallel neural circuits, many of which are associated with social emotions, evaluate the actions and intentions of others. Through various mechanisms of competition, only a subset of these circuits ultimately causes a decision or an action. This activity is experienced consciously as a subjective moral sense of right or wrong, and an interpretive process offers *post hoc* explanations designed to link the social stimulus with the subjective moral response using whatever explicit information is available.

Address
SAGE Center for the Study of Mind, University of California, Santa Barbara, Santa Barbara, CA 93106-9660, United States

Corresponding author: Funk, Chadd M (cmfunk@wisc.edu) and Gazzaniga, Michael S (M.Gazzaniga@psych.ucsb.edu)

Current Opinion in Neurobiology 2009, **19**:678–681

This review comes from a themed issue on
Neurobiology of behaviour
Edited by Catherine Dulac and Giacomo Rizzolatti

Available online 4th November 2009

0959-4388/$ – see front matter

DOI 10.1016/j.conb.2009.09.011

Introduction and background

The first step toward understanding the neural basis of human morality is acknowledging that it is embodied in, and thus operates based on the principles of, the functional architecture of the human brain, a complex system consisting of a wide array of neural circuits that operate as functional modules and are selectively engaged by environmental demands [1]. The circuits are organized in decentralized, highly parallel fashion, such that a large amount of information is processed simultaneously in an ongoing effort to produce adaptive behavior [2]. The requirement of producing serial adaptive action is a powerful biological constraint. In order for parallel circuits to generate serial functionality in the absence of a central controller, there must be competition for limited resources that enable the sort of amplification of activity within a circuit or set of related-circuits necessary to bias the competition. In the case of moral processing, we must seek to characterize the subset of modules that are engaged by moral stimuli and consider how the interactions between these modules result in adaptive behavior. Such an explanation would account for the causal component of moral judgments.

In the following sections, we review recent research that provides new clues about how the brain accomplishes moral tasks. We begin by considering the type of actions that morality promotes, as this is the critical output that must have been adaptive in order for morality to emerge. We then review the neural circuits activated by various moral judgment paradigms and consider possible mechanisms of interaction between opposing modules. Next, we address evidence for an interpretive process that attributes subjective moral feelings to specific stimuli and generates an explanation for the link between the two. This interpretive process uses only the end result of the cacophony of parallel processing and competition, enabling a coherent, compelling moral narrative to develop. The narratives generated by this process form the explicit norms that are recorded and propagated by various societal institutions, which provide important feedback that fine-tunes our moral neural circuitry and that of future generations.

Producing adaptive behavior in a social context

The central purpose of the human brain is to produce adaptive responses to whatever it encounters in the environment. For humans, the environment imposes both physical and social demands. Generating adaptive behavior in response to the latter is a special challenge and may have been a driving force behind the emergence of human morality [3•,4]. In brief, it is necessary to ascertain with whom we should interact and possibly invest the resources required to maintain relationships. In order to solve this problem, the brain constantly evaluates whether to approach or avoid other members of the species [5]. Social evaluation is such a fundamental capacity that it appears to be hard-wired into the infant brain. Premack and Premack [6] reported that infants understood that self-propelled, goal-directed objects possess intentions, and furthermore, the infants had a basic understanding of positive and negative valence across different actions. A recent study replicated this finding and extended it by showing that preverbal babies, aged 6 and 10 months, have a basic understanding of helping and hurting, and moreover, that they will choose to play with

the helper when given a choice [7•]. Simply put, social evaluation impacts subsequent action. Human morality is perhaps rooted in this social evaluation.

Such evaluation is possibly an extension of more primitive systems that lead to other very basic approach or avoid behaviors, such as food rejection [8,9•]. In these systems, emotions (such as disgust, fear, or pleasure) provide the impetus for avoiding or approaching. In the same way, social evaluation likely depends on social emotions. Therefore, if human morality is rooted in social evaluation, then we might expect to find that social emotions play a fundamental role in moral processing. In the next section, we review evidence that this is precisely the case.

Neural circuitry of moral processing

The pioneering work of Greene *et al.* [10], the first to present subjects with classic moral dilemmas in an fMRI setting, provided compelling support for the hypothesis that emotions played an integral role in moral judgments. When subjects were given moral dilemmas that involved personal actions (such as pushing an overweight person off a footbridge to stop a trolley headed for five people), emotional regions of interest, including the medial frontal gyrus, posterior cingulate gyrus, and angular gyrus were more active relative to when subjects were given dilemmas featuring impersonal actions (such as flipping a switch that would change the path of the train, saving the five people but killing one). Impersonal dilemmas selectively engaged areas associated with working memory, including regions of middle frontal gyrus and posterior parietal cortex. Presumably, these regions support the abstract reasoning necessary to weigh the benefits of the possible courses of action.

More recently, the critical role that automatic emotional processing plays in moral function has been solidified [11]. Moral disgust engages widely distributed emotional brain areas, many of which overlap with regions active during pathogen-induced disgust [12]. Inequity aversion in a distributive justice paradigm, in which participants made real-life meal allocation decisions, was associated with increased activity in the insula [13••]. On the other end of the social emotion spectrum, admiration and compassion were associated with emotional networks in the anterior cingulate, anterior insula, and hypothalamus, as well as subregions of the posteromedial cortices [14••]. Thus, both avoid and approach-related social emotions are vital to moral evaluation.

But this is not to say that brain networks underlying social emotions are the only neural circuits relevant to moral decision-making. For instance, other neural circuits function to represent and evaluate the intentions of others. This is of particular importance when an intended action and the result actually achieved differ, such as when an actor possesses the intention to harm but does not cause the harm upon executing the planned act. Recent studies indicate that activity in a network implicated in belief attribution, prominently featuring the right temporoparietal junction (RTPJ), is selectively elevated when subjects encounter negative belief information [15•,16] or spontaneously infer belief information while judging situations with negative outcomes [17].

Interestingly, split-brain patients do not judge moral violations on the basis of an agent's mental state when belief and outcome are inconsistent; instead, they rely entirely on outcome (Miller *et al.*, unpublished data). This peculiar finding may be explained by the fact that the left hemisphere, the hemisphere responding verbally to the dilemmas, does not receive input from key nodes in the aforementioned belief-attribution network (such as the RTPJ). Moreover, separate evidence indicates that the right inferior frontal cortex is specialized for modeling the intentions of others via a mirror neuron mechanism [18], which would further imply that the disconnected left hemisphere possesses an extremely limited ability to integrate intention-based information into moral judgments. These findings illustrate that neural circuits not directly related to social emotions make important contributions to normal moral processing.

In essence, the research reviewed in this section identifies important anatomical circuits that are activated by specific types of social stimuli. However, most dilemmas or decisions in the social world elicit parallel activity in many of these circuits, and this activity may simultaneously bias motor systems toward opposite avoid/approach behaviors. Different types of moral dilemmas elicit unique patterns of neural response [19], and emotional circuits do not always dictate behavior. In order to understand how adaptive behavior emerges, it is necessary to consider interactions between circuits.

Competition yields adaptive behavior

In a decentralized architecture, there is no central authority to make decisions or select actions. Instead, competition for limited resources resolves the conflict, resulting in amplified activity within the dominant subset of modules that then biases decision-making or action selection.

The difficulty of many moral dilemmas is an extreme manifestation of this competition. Difficult personal moral dilemmas, when contrasted with easy personal moral dilemmas, induce conflict-related activity in the anterior cingulate cortex (ACC) and subsequent control-related activity in the anterior dorsolateral prefrontal cortex (DLPFC) [20]. Furthermore, across only difficult trials, the DLPFC was even more active when individuals made utilitarian judgments relative to when they made nonutilitarian judgments, suggesting that the dampening activity of this region is required to overcome the automatic emotional response to the dilemma. When DLPFC

modules are occupied with other tasks, utilitarian representations compete less efficiently with emotional responses and reaction time increases [21].

It is worth restating that this is an extreme example of competition. In fact, competition between neural circuits likely exists on a continuum. The other end of the continuum is illustrated by patients with ventromedial prefrontal cortex damage. These patients endorse utilitarian action in difficult personal dilemmas at a significantly higher rate than controls [22•], as if there were no emotional consideration to compete with the utilitarian appraisal of these dilemmas. Somewhere between these extremes, it is likely that there is competition that is easily resolved and thus does not require extensive recruitment of ACC and DLPFC regions. For instance, when easy personal dilemmas induce a strong emotional response, emotional activity outcompetes cognitive appraisals. Or representations of actions may be outcompeted by representations of intentions, which in turn dictate moral valence. In these cases, the limited resources that constrain amplification of activity to a subset of neural circuits (a phenomenon that is not well understood) may be sufficient to govern competition. As competition ramps up, other circuits, such as frontal regions involved in conflict detection and cognitive control [20], may be activated to inhibit activity in certain circuits, ultimately influencing a decision or impacting an action by biasing competition.

The role of interpretation in moral thinking

Up until now, we have focused on the neural circuits and interactions that underlie moral judgment. Activity in these circuits also manifests in conscious content, shaping our felt states and endowing our subjective experience with a dimension of right and wrong. But what is the utility of our subjective moral sense if the causal work is already done? Humans possess a strong conviction that deliberate moral reasoning leads to judgments, but the evidence reviewed thus far casts substantial doubt on this idea. We suggest that, rather than a causal determinant in the moral decision-making process, moral reasoning is most usefully thought of as an attempt to explain the cause and effect of our moral intuitions that draws upon all available explicit information about a given situation.

We posit that an interpretive process, localized in the left cerebral hemisphere, underlies this explanatory drive and is responsible for the moral hypotheses that link social stimuli to felt states in a coherent way. The interpreter was originally discovered when asking the left hemisphere of split-brain subjects to describe behavior induced by the right hemisphere in response to stimuli that the left hemisphere was not privy to [23]. The left hemisphere used information available to it in order to generate a seemingly reasonable explanation. Recently, the activity of the interpreter has been observed in a moral judgment paradigm (Miller *et al.*, unpublished data). In the split-brain experiment described previously, the left hemisphere made judgments solely based on the outcome of an action and not based on mental state, perhaps due to a loss of input from important belief-attribution [15•] and intention-modeling [18] networks in the right hemisphere. When asked to explain why she rated innocent acts that accidently caused harm as forbidden, one patient provided complicated explanations in an attempt to justify her judgment.

Similarly, by hypnotizing healthy subjects to feel disgust at neutral words, Wheatley and Haidt [24] were able to manipulate the subjects' moral judgments. Upon hearing a story involving a class president organizing student discussions with faculty (that featured the disgust-inducing neutral words), some subjects condemned the class president's actions. When asked to explain this rating, subjects stated that they were suspicious of the class president or provided other creative reasons for their judgment. This behavior suggests that automatic social evaluation produces a judgment, which the interpreter registers and attempts to explain.

The interpreter introduces a misleading level of certainty about the reasons that moral judgments are made. Yet the narrative of the interpreter helps us make sense of our social environment. It provides a critical bridge from the undeniable subjective elements of our ongoing conscious experience to the explicit ideas and convictions that, via communication, eventually crystallize to form the ideological infrastructure of society. Once captured as cultural norms or laws, these ideas feedback through development and learning mechanisms to fine-tune the workings of the underlying neural circuitry [25]. Indeed, recent findings indicate that cultural influences have a substantial effect on cognitive processes [26,27], including moral processing (Hauser, unpublished data).

Conclusion

Thus, hard-wired patterns of neural connectivity that establish innate functional modules, like those that foster basic social evaluation in infants, are dynamically sculpted by cultural experience. On the basis of the findings reviewed above and the principles of decentralized parallel processing and competition, there is support for a tentative model for individual differences in moral convictions rooted in the sensitivities of the various circuits described earlier. For instance, if genetic factors and activity-dependent processes throughout development strengthen connections within a neural circuit sensitive to inequity, such as the one prominently involving the insula [13••], then activation of this subnetwork may result in especially robust activity that outcompetes activity in other modules and leads to a certain bias in moral judgments, in this case an equity bias. Individual

differences in opinion on moral topics within a given society may be based on the sensitivities of specific neural circuits that process various moral dimensions [3•]. Future research should attempt to characterize the neural properties that lead to particularly effective levels and patterns of activity in these circuits. Such research would also illuminate mechanisms of competition between functional modules, which would ultimately lead to a more complete understanding of how our moral brains operate.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Gazzaniga MS: **Organization of the human brain**. *Science* 1989, **245**:947-952.

2. Gazzaniga MS, Doron KW, Funk CM: **Looking toward the future: perspectives on examining the architecture and function of the human brain as a complex system**. In *The Cognitive Neurosciences,* edn 4. Edited by Gazzaniga MS. MIT Press; 2009:1247-1254.

3. • Haidt J: **The new synthesis in moral psychology**. *Science* 2007, **316**:998-1002.
This paper identifies important undercurrents in the growing scientific literature on human morality, spanning psychology, evolutionary psychology, and neurobiology.

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

5. Gazzaniga MS: *Human: The Science Behind What Makes us Unique* New York: HarperCollins; 2008.

6. Premack D, Premack AJ: **Infants attribute value ± to the goal-directed actions of self-propelled objects**. *J Cogn Neurosci* 1997, **9**:848-856.

7. • Hamlin JK, Wynn K, Bloom P: **Social evaluation by preverbal infants**. *Nature* 2007, **450**:557-559.
This work demonstrates that preverbal infants assign valence to helping and hurting behaviors and subsequently base decisions about which actor to play with on this information.

8. Haidt J, Rozin P, McCauley C, Imada S: **Body, psyche, culture: the relationship between disgust and morality**. *Psychol Dev Societies* 1997, **9**:107-131.

9. • Chapman HA, Kim DA, Susskind JM, Anderson AK: **In bad taste: evidence for the oral origins of moral disgust**. *Science* 2009, **323**:1222-1226.
This research demonstrates that moral disgust activated the same facial muscles that were active when subjects sampled unpleasant tasting liquids or viewed pictures of unclean or contaminated objects.

10. Greene JD, Sommerville RB, Nystrom LE, Darley JM, Cohen JD: **An fMRI investigation of emotional engagement in moral judgment**. *Science* 2001, **293**:2105-2108.

11. Greene J, Haidt J: **How (and where) does moral judgment work?** *Trends Cogn Sci* 2002, **6**:517-523.

12. Borg JS, Lieberman D, Kiehl KA: **Infection, incest, and iniquity: investigating the neural correlates of disgust and morality**. *J Cogn Neurosci* 2008, **20**:1529-1546.

13. •• Hsu M, Anen C, Quartz SR: **The right and the good: distributive justice and neural encoding of equity and efficiency**. *Science* 2008, **320**:1092-1095.
This work shows that increases in insula activity were associated with heightened sensitivity to inequity in a distributive justice paradigm that involved making real-life decisions.

14. •• Immordino-Yang MH, McColl A, Damasio H, Damasio A: **Neural correlates of admiration and compassion**. *Proc Natl Acad Sci U S A* 2009, **106**:8021-8026.
This study features a novel experimental paradigm designed to induce positive moral emotions of compassion and admiration in an fMRI setting. The authors report recruitment of many structures associated with emotional function.

15. • 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.
This paper reports that activity in RTPJ increased in all conditions in which belief information was presented and that RTPJ activity was further increased when negative beliefs were followed by neutral outcomes.

16. Young L, Saxe R: **The neural basis of belief encoding and integration in moral judgment**. *Neuroimage* 2008, **40**:1912-1920.

17. Young L, Saxe R: **An FMRI investigation of spontaneous mental state inference for moral judgment**. *J Cogn Neurosci* 2009, **21**:1396-1405.

18. Iacoboni M, Molnar-Szakacs I, Gallese V, Buccino G, Mazziotta JC, Rizzolatti G: **Grasping the intentions of others with one's own mirror neuron system**. *PLoS Biol* 2005, **3**:e79.

19. Schaich Borg J, Hynes C, Van Horn J, Grafton S, Sinnott-Armstrong W: **Consequences, action, and intention as factors in moral judgments: an FMRI investigation**. *J Cogn Neurosci* 2006, **18**:803-817.

20. 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.

21. Greene JD, Morelli SA, Lowenberg K, Nystrom LE, Cohen JD: **Cognitive load selectively interferes with utilitarian moral judgment**. *Cognition* 2008, **107**:1144-1154.

22. • Koenigs M, Young L, Adolphs R, Tranel D, Cushman F, Hauser M, Damasio A: **Damage to the prefrontal cortex increases utilitarian moral judgements**. *Nature* 2007, **446**:908-911.
This study found that VMPFC patients were significantly more likely to endorse utilitarian responses to difficult personal moral dilemmas than controls.

23. Gazzaniga MS, LeDoux JE: *The Integrated Mind*. New York: Plenum Press; 1978.

24. Wheatley T, Haidt J: **Hypnotic disgust makes moral judgments more severe**. *Psychol Sci* 2005, **16**:780-784.

25. Flack JC, Krakauer DC: **Evolution and construction of moral systems**. In *Games, Groups, and the Global Good.* Edited by Levin SA. Springer; 2009:117-141.

26. Masuda T, Nisbett RE: **Attending holistically versus analytically: comparing the context sensitivity of Japanese and Americans**. *J Pers Soc Psychol* 2001, **81**:922-934.

27. Uskul AK, Kitayama S, Nisbett RE: **Ecocultural basis of cognition: farmers and fishermen are more holistic than herders**. *Proc Natl Acad Sci U S A* 2008, **105**:8552-8556.