# The lateral prefrontal cortex and moral goal pursuit

Ryan W Carlson and Molly J Crockett

In the face of competing desires, humans often strive to be fair, honest, and considerate of others. Research from social neuroscience implicates the lateral prefrontal cortex (LPFC) in our capacity to pursue such goals, yet its precise computational role is less clear. Here, we draw on insights from the neuroscience of hierarchical control and value-based choice to offer an integrative look at how LPFC supports the pursuit of moral goals. We conclude by highlighting how future work may leverage these insights to deepen our understanding of the dynamic neural code of morality.

### Address
Department of Psychology, Yale University, USA

Corresponding author: Crockett, Molly J (molly.crockett@yale.edu)

Humans routinely forgo immediate desires in order to be fair, honest, and respectful. Such acts highlight our species' capacity both to value *moral norms*—prescriptive and proscriptive rules surrounding ethical conduct [1]—and to actively pursue *goal-directed* actions to uphold these norms.

Theories and evidence from social neuroscience suggest that our ability to pursue goals of this nature—or *moral goals*[1]—is often guided by the lateral prefrontal cortex (LPFC) [2**,3–6]. Though despite its central place in models of how the brain coordinates moral behavior, the rich computational architecture of this structure is often overlooked.

In this brief review, we highlight recent insights that illuminate how LPFC's architecture both *represents* moral goals (via hierarchical control processes) and *guides* moral actions (via value-based processes). Together, these insights offer a richer, *goal-based* view of how LPFC guides moral behavior.

## The computational architecture of LPFC
LPFC is implicated in a wide range of functions in human cognition [7], though perhaps its most central function is to actively represent our goals and strategies for pursuing them [8–10]. LPFC broadly comprises rostrolateral (RLPFC), ventrolateral (VLPFC), and dorsolateral (DLPFC) subregions (See Figure 1)—each of which play complementary roles in representing and maintaining goal pursuit [11]. Importantly, their roles can be functionally distinguished by at least three features of LPFC's architecture: (i) hierarchical organization, (ii) connectivity patterns, and (iii) lateralization.

### Hierarchical organization
Social goals are usually structured *hierarchically*, such that a high-level goal (e.g., to make a good impression during a job interview) is composed of more concrete, subgoals (e.g., executing a respectable handshake). Crucially, models of hierarchical control suggest that LPFC is structured to represent goals in precisely this way. Indeed, robust evidence suggests that LPFC is hierarchically organized along a roughly rostral-caudal axis [12,13,14**], wherein more *abstract* representations pertaining to high-level goals are encoded more rostrally, and more *concrete* representations related to actions (and subgoals) are encoded more caudally. Recent work proposes that mid-DLPFC sits atop this hierarchy[2] [12], wherein more rostral and caudal inputs converge with *contextual* representations—thus allowing competing goals to be flexibly pursued based on their appropriateness for the situation. In support of this view, RLPFC is typically sensitive to future-oriented aspects of goal pursuit, such as planning future actions and monitoring goal progress [15]. In contrast, more caudal areas of VLPFC and DLPFC—extending into the premotor cortex— are often engaged while actively controlling motor responses to one's immediate environment [11].

### Connectivity
LPFC's ability to implement goals depends on reciprocal interactions with a widely distributed set of brain regions—including core structures for encoding value (e.g., ventromedial prefrontal cortex [VMPFC] and

---

[1] Here, a *goal* denotes a structured, internal representation of a desired state. A *moral goal* is one directed towards a desired state that aligns with a moral norm (e.g., a fair distribution of a resource). Individuals typically value such states intrinsically—*as an end* in itself—but may also value these states instrumentally—*as a means* to other goals (e.g., avoiding punishment or guilt).

[2] That is, whereas RLPFC is positioned at the top of the *representational hierarchy* in LPFC, mid-DLPFC is at the top of the *control hierarchy*—exerting an asymmetrically greater influence over other LPFC subregions [14**].

**Figure 1**
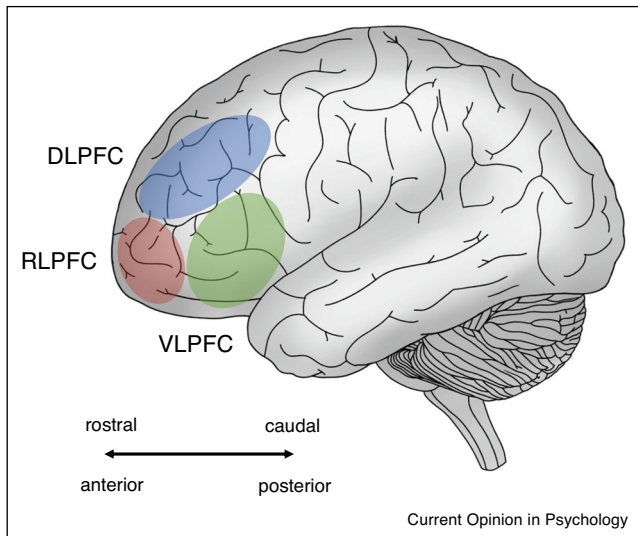


Approximate locations of the broadest subdivisions of LPFC [74]. *Abbreviations*: RLPFC, rostrolateral prefrontal cortex (area 10); VLPFC, ventrolateral prefrontal cortex (areas 47/45/44); DLPFC, dorsolateral prefrontal cortex (areas 9/46).

striatum), affect (e.g., anterior insula [AI]), and mental states (e.g., posterior cingulate cortex [PCC]), as well as those that *initiate* goals through conflict detection and goal selection mechanisms (e.g., anterior cingulate cortex) [11,16–18].

### Lateralization

Additionally, some functions tend to be lateralized to one hemisphere of LPFC [19]. For instance, the ability to suppress prepotent responses—or *response inhibition*—has been localized to more caudal areas of the right VLPFC [20]. In addition, disruption of the right (more so than left) DLPFC impairs people's ability to pursue goals based on internal models of their environment [21].

## A flexible role for LPFC in moral goal pursuit

A distinct yet complementary line of research in social neuroscience also implicates LPFC in our propensity to uphold moral norms. Typically, these experiments use iterated social dilemmas that pit motives to uphold a moral norm against motives to maximize one's financial reward. Here, we focus on studies of two representative moral norms: fairness and honesty.

### Fairness

Numerous studies find that activation of the right DLPFC increases as individuals more fairly distribute resources with a social partner, both when such norms are made salient by a partner's expectations [22,23], or the threat of being punished by them [24]. Interestingly, deciding under the threat of punishment (compared to

when no threat exists) also engages RLPFC [24], potentially reflecting a goal-directed increase in planning (and monitoring) of one's long-term social strategy.

Studies leveraging *non-invasive brain stimulation* (NIBS) offer more direct evidence of LPFC's role in norm-guided behavior. Such studies often find that decreasing the excitability of neurons in the right DLPFC *reduces* people's level of fairness toward a social partner—thus damaging their reputation and future prospects [25–28]. Crucially, while inhibiting right DLPFC causes individuals to act less fairly, they nonetheless retain explicit knowledge about how fair their behavior was [25,26]—supporting the view that such representations are maintained more rostrally in LPFC [75].

In contrast, when people are deciding in contexts in which moral norms are less salient and well-defined (i.e., splitting money *without* the risk of future interaction, feedback, or sanctions from a partner), the relationship between LPFC and moral behavior sometimes reverses. In such cases, DLPFC activation has been associated with more selfish strategies [26,29,30]. These studies highlight *flexibility* in the types of goals LPFC implements, and how the *activation* and *intensity* of moral versus selfish goals can shift across contexts.

### Honesty

LPFC engagement is also associated with choosing to behave honestly when individuals are tempted by dishonesty. For instance, in tasks where dishonest responses can yield financial gains, those prone to dishonesty show increased DLPFC and VLPFC activation when behaving honestly [31], and increasing the excitability of right DLPFC via NIBS among such groups increases their tendency to respond honestly [32]. In addition, DLPFC lesion patients appear to be unaffected by honesty norms when making choices [33].

Other work further suggests that the role of LPFC in upholding moral norms depends on the *relative* intensity of conflicting moral and monetary desires. For instance, higher sensitivity to financial rewards—indexed by ventral striatal activity—predicts the rate to which individuals violate honesty norms in order to gain money [34], and also predicts DLPFC engagement when such individuals *forgo* money to behave honestly. Other findings highlight similar interactions between LPFC and other value-sensitive regions. For instance, increased engagement of VMPFC, as well as increased connectivity between VMPFC and DLPFC, both predict dishonest behavior—independent of whether dishonesty benefitted the self or someone else [35]. Such work converges with other recent studies on honesty [36], and prosociality more broadly [37,38], which highlight key individual and contextual differences in how VMPFC-DLPFC coupling relates to moral, versus selfish, behavior.

Overall, these findings highlight a reliable, yet *flexible*, role for LPFC in guiding moral behavior. Although LPFC function is frequently linked with moral behavior, LPFC may ultimately serve to promote either selfish or moral goals across different contexts. Such flexibility is consistent with how this structure is viewed from the lens of hierarchical control [14••]. It also reflects another feature of goal pursuit we have alluded to: pursuing goals often involves dynamically updating the *value* of different actions and outcomes within one's social environment.

## LPFC aligns value representations with active goals

Early work often inferred that DLPFC activity observed during moral behavior reflected this structure *inhibiting* prepotent selfish responses [4]. However, such proposals have encountered a number of challenges [2••]. For instance, recent work supports the view that more concrete strategies such as response inhibition are not implemented by DLPFC, but instead by more 'low-level' caudal areas in the right VLPFC [20,39]. Moreover, growing research on value-based choice has now mapped a broader and more flexible set of functions across LPFC's hierarchical architecture that could support moral goals [2,11,40••,41,42]. Most notably, a recent proposal [2••] instead suggests that DLPFC activity observed during moral behavior may reflect this structure *updating* the value associated with desirable actions that are contextually incompatible with a moral goal.

Growing evidence supports and extends this proposal, suggesting that structures across LPFC can regulate goal pursuit via dynamic interactions with other value-sensitive regions [11,18,43–45]. For instance, through functional loops with VMPFC, LPFC can flexibly integrate the value of goal-relevant features—such as abstract rules [11] and future states [46,47]—to guide choice. Moreover, LPFC directly regulates basic action value representations via projections to the striatum—a key region involved in action selection [48–50]. These findings imply that LPFC could dampen the desirability of an immediately gratifying action (e.g., eating your roommate's mint chip ice cream) in the service of a long-term goal (e.g., maintaining a healthy relationship with your roommate).

A recent study directly tested the idea that LPFC integrates moral norms into action values to guide choice [51••]. This study focused on the moral norm against physically harming others. Specifically, individuals made choices involving a conflict between gaining money for themselves and inflicting pain on either themselves or a partner. Most individuals' choices indicated sensitivity to norms against harming others, as they required more money to inflict pain on others than themselves [52,53]. Moreover, these preferences tracked with neural responses to monetary gains, such that those with stronger moral preferences showed *decreased* responses in a network of value-sensitive regions (including the dorsal striatum [DS], AI, PCC, and VMPFC) to money gained by harming others compared to oneself. In other words, ill-gotten gains activated the brain's value circuitry less strongly than money earned decently. Crucially, a cluster of LPFC activity—bordering the left DLPFC and left VLPFC—tracked with the blameworthiness of harmful choices, and connectivity between LPFC and DS became more negative when participants made helpful choices.

LPFC was not, however, more active during helpful (versus harmful) choices, nor when helpful choices were more difficult. Activity patterns in LPFC instead reflected the encoding of moral goals, which in turn regulated action values represented in the striatum. Overall, these findings align with value-based models, wherein LPFC can flexibly guide behavior by updating the value of otherwise desirable actions that conflict with active goals.

More broadly, echoing other recent work [54–58], these data highlight how the mechanisms that guide moral behavior depend on interactions between *individuals* (e.g., the relative value they ascribe to moral norms versus competing rewards) and *contexts* (e.g., to what extent moral norms are salient, well-defined, or instrumental to other social and emotional goals). As such, the role of LPFC in moral goal pursuit is best viewed as reliable, but flexible. Its architecture is well-equipped to resolve the computational problem of aligning one's actions with moral norms in novel, distracting, and complex social settings—but individuals must first be motivated as such. When immediate desires do conflict with active moral goals, growing interdisciplinary evidence supports the view that LPFC acts as a sophisticated and flexible guide—situating the value of potential actions within a moral context.

## Future directions
The work reviewed above primarily examined instances of norm *compliance*; however, LPFC likely also supports goals to *enforce* moral norms. Theoretical work supports the possibility that norm compliance and enforcement share common psychological mechanisms [59,60]. Moreover, in addition to being engaged when individuals uphold moral norms, LPFC is similarly engaged when individuals enforce them. For instance, increased activity in right DLPFC is observed when people incur a cost to punish a partner who unfairly allocates money [61,62], and when observers judge an agent's culpability for a criminal act [63]. Moreover, in much the same way applying NIBS to LPFC alters norm compliance, numerous studies find that disrupting right DLPFC reduces peoples' tendency to disapprove of [64], and punish [61,65••], people who violate fairness norms. Together, these studies indicate a notable degree of neural overlap

between decisions to enforce, and comply with, moral norms. Indeed, assessing whether LPFC plays a similar role in pursuing moral goals via compliance and enforcement remains an important direction for future work to explore.

Additionally, most studies using NIBS have targeted the DLPFC. However, the coarse spatial resolution of NIBS [66], as well as the deeply interconnected architecture of LPFC, may hinder the ability of such techniques to test how different goal pursuit strategies are represented across LPFC. Combining NIBS and computational neuroscience techniques, as well as modelling the unique temporal profiles of LPFC's subregions [67], may prove fruitful for constraining predictions and inferences in future work.

Finally, future studies should further explore how people adjudicate between conflicting moral goals. Such conflicts can arise between empathy (maximizing the welfare of one victim) and fairness (maximizing the welfare of the group) [68], but also efficiency versus equity [69], and consequentialism versus non-consequentialism [70,71]. LPFC function is often studied within the context of deciding between a concrete, immediately gratifying, and typically 'suboptimal' action, versus an abstract, delayed, 'optimal' action. However, individuals face many other types of moral dilemmas in social life [72], and can adopt numerous strategies to overcome them [73]. More closely examining how the brain resolves such dilemmas will no doubt yield richer insights into how, and when, LPFC contributes to human morality.

## Conflict of interest statement
Nothing declared.

## Acknowledgements

## References

1. Turiel E: *The Development of Social Knowledge: Morality and Convention*. Cambridge University Press; 1983.

2. Buckholtz JW: **Social norms, self-control, and the value of
•• antisocial behavior**. *Curr Opin Behav Sci* 2015, **3**:122-129.
This seminal paper 'recasts' norm compliance as a value-based decision problem, instead of a response-inhibition problem, and provides a compelling case that DLPFC can integrate norms into action values to promote norm-guided behavior.

3. Buckholtz JW, Marois R: **The roots of modern justice: cognitive and neural foundations of social norms and their enforcement**. *Nat Neurosci* 2012, **15**:655.

4. Knoch D, Fehr E: **Resisting the power of temptations**. *Ann N Y Acad Sci* 2007, **1104**:123-134.

5. Moll J, de Oliveira-Souza R: **Moral judgments, emotions and the utilitarian brain**. *Trends Cogn Sci* 2007, **11**:319-321.

6. 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.

7. Fuster J: *The Prefrontal Cortex*. Academic Press; 2015.

8. Miller EK, Cohen JD: **An integrative theory of prefrontal cortex function**. *Annu Rev Neurosci* 2001, **24**:167-202.

9. Gerlach KD, Spreng RN, Gilmore AW, Schacter DL: **Solving future problems: default network and executive activity associated with goal-directed mental simulations**. *NeuroImage* 2011, **55**:1816-1824.

10. Asplund CL, Todd JJ, Snyder AP, Marois R: **A central role for the lateral prefrontal cortex in goal-directed and stimulus-driven attention**. *Nat Neurosci* 2010, **13**:507-512.

11. Dixon ML, Christoff K: **The lateral prefrontal cortex and complex value-based learning and decision making**. *Neurosci Biobehav Rev* 2014, **45**:9-18.

12. Badre D, Nee DE: **Frontal cortex and the hierarchical control of behavior**. *Trends Cogn Sci* 2018, **22**:170-188.

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

14. Nee DE, D'Esposito M: **The hierarchical organization of the
•• lateral prefrontal cortex**. *eLife* 2016, **5**:e12112.
This recent study leverages dynamic causal modelling to elucidate the computational architecture of LPFC and its contributions to hierarchical control.

15. Desrochers TM, Chatham CH, Badre D: **The necessity of rostrolateral prefrontal cortex for higher-level sequential behavior**. *Neuron* 2015, **87**:1357-1368.

16. Alexander WH, Brown JW: **Hierarchical error representation: a computational model of anterior cingulate and dorsolateral prefrontal cortex**. *Neural Comput* 2015, **27**:2354-2410.

17. Duverne S, Koechlin E: **Rewards and cognitive control in the human prefrontal cortex**. *Cereb Cortex* 2017, **27**:5024-5039.

18. Shenhav A, Botvinick MM, Cohen JD: **The expected value of control: an integrative theory of anterior cingulate cortex function**. *Neuron* 2013, **79**:217-240.

19. Shallice T, Cipolotti L: **The prefrontal cortex and neurological impairments of active thought**. *Annu Rev Psychol* 2018, **69**:157-180.

20. Aron AR, Robbins TW, Poldrack RA: **Inhibition and the right inferior frontal cortex: one decade on**. *Trends Cogn Sci* 2014, **18**:177-185.

21. Smittenaar P, FitzGerald THB, Romei V, Wright ND, Dolan RJ: **Disruption of dorsolateral prefrontal cortex decreases model-based in favor of model-free control in humans**. *Neuron* 2013, **80**:914-919.

22. Chang LJ, Smith A, Dufwenberg M, Sanfey AG: **Triangulating the neural, psychological, and economic bases of guilt aversion**. *Neuron* 2011, **70**:560-572.

23. Nihonsugi T, Ihara A, Haruno M: **Selective increase of intention-based economic decisions by noninvasive brain stimulation to the dorsolateral prefrontal cortex**. *J Neurosci* 2015, **35**:3412-3419.

24. Spitzer M, Fischbacher U, Herrnberger B, Grön G, Fehr E: **The neural signature of social norm compliance**. *Neuron* 2007, **56**:185-196.

25. Knoch D, Schneider F, Schunk D, Hohmann M, Fehr E: **Disrupting the prefrontal cortex diminishes the human ability to build a good reputation**. *Proc Natl Acad Sci* 2009, **106**:20895-20899.

26. Ruff CC, Ugazio G, Fehr E: **Changing social norm compliance with noninvasive brain stimulation**. *Science* 2013, **342**:482-484.

27. Soutschek A, Sauter M, Schubert T: **The importance of the lateral prefrontal cortex for strategic decision making in the prisoner's dilemma**. *Cogn Affect Behav Neurosci* 2015, **15**:854-860.

28. Strang S, Gross J, Schuhmann T, Riedl A, Weber B, Sack AT: **Be nice if you have to — the neurobiological roots of strategic fairness**. *Soc Cogn Affect Neurosci* 2015, **10**:790-796.

29. FeldmanHall O, Dalgleish T, Thompson R, Evans D, Schweizer S, Mobbs D: **Differential neural circuitry and self-interest in real vs hypothetical moral decisions**. *Soc Cogn Affect Neurosci* 2012, **7**:743-751.

30. Yamagishi T, Takagishi H, Fermin ADSR, Kanai R, Li Y, Matsumoto Y: **Cortical thickness of the dorsolateral prefrontal cortex predicts strategic choices in economic games**. *Proc Natl Acad Sci* 2016, **113**:5582-5587.

31. Greene JD, Paxton JM: **Patterns of neural activity associated with honest and dishonest moral decisions**. *Proc Natl Acad Sci* 2009, **106**:12506-12511.

32. Maréchal MA, Cohn A, Ugazio G, Ruff CC: **Increasing honesty in humans with noninvasive brain stimulation**. *Proc Natl Acad Sci* 2017, **114**:4360-4364.

33. Zhu L, Jenkins AC, Set E, Scabini D, Knight RT, Chiu PH, King-Casas B, Hsu M: **Damage to dorsolateral prefrontal cortex affects tradeoffs between honesty and self-interest**. *Nat Neurosci* 2014, **17**:1319.

34. Abe N, Greene JD: **Response to anticipated reward in the nucleus accumbens predicts behavior in an independent test of honesty**. *J Neurosci* 2014, **34**:10564-10572.

35. Pornpattananangkul N, Zhen S, Yu R: **Common and distinct neural correlates of self-serving and prosocial dishonesty**. *Hum Brain Mapp* 2018, **39**:3086-3103.

36. Dogan A, Morishima Y, Heise F, Tanner C, Gibson R, Wagner AF, Tobler PN: **Prefrontal connections express individual differences in intrinsic resistance to trading off honesty values against economic benefits**. *Sci Rep* 2016, **6**:33263.

37. Hu J, Li Y, Yin Y, Blue PR, Yu H, Zhou X: **How do self-interest and other-need interact in the brain to determine altruistic behavior?** *NeuroImage* 2017, **157**:598-611.

38. Wills J, Hackel L, Bavel JJV: **Shifting prosocial intuitions: neurocognitive evidence for a value based account of group-based cooperation**. *PsyArXiv* 2018 http://dx.doi.org/10.31234/osf.io/u736d. preprints.

39. Stramaccia DF, Penolazzi B, Sartori G, Braga M, Mondini S, Galfano G: **Assessing the effects of tDCS over a delayed response inhibition task by targeting the right inferior frontal gyrus and right dorsolateral prefrontal cortex**. *Exp Brain Res* 2015, **233**:2283-2290.

40. O'Doherty JP, Cockburn J, Pauli WM: **Learning, reward, and
•• decision making**. *Annu Rev Psychol* 2017, **68**:73-100.
An accessible overview of the neuroscience of reinforcement learning and decision-making which highlights the interactive nature of the brain's goal-directed and habitual systems.

41. Ruff CC, Fehr E: **The neurobiology of rewards and values in social decision making**. *Nat Rev Neurosci* 2014, **15**:549.

42. Crockett MJ, Braams BR, Clark L, Tobler PN, Robbins TW, Kalenscher T: **Restricting temptations: neural mechanisms of precommitment**. *Neuron* 2013, **79**:391-401.

43. Etkin A, Büchel C, Gross JJ: **The neural bases of emotion regulation**. *Nat Rev Neurosci* 2015, **16**:693.

44. Tusche A, Hutcherson CA: **Cognitive regulation alters social and dietary choice by changing attribute representations in domain-general and domain-specific brain circuits**. *eLife* 2018, **7**:e31185.

45. Mansouri FA, Koechlin E, Rosa MGP, Buckley MJ: **Managing competing goals — a key role for the frontopolar cortex**. *Nat Rev Neurosci* 2017, **18**:645-657.

46. Hare TA, Hakimi S, Rangel A: **Activity in dlPFC and its effective connectivity to vmPFC are associated with temporal discounting**. *Front Neurosci* 2014, **8**:50.

47. Rudorf S, Hare TA: **Interactions between dorsolateral and ventromedial prefrontal cortex underlie context-dependent stimulus valuation in goal-directed choice**. *J Neurosci* 2014, **34**:15988-15996.

48. Deserno L, Huys QJ, Boehme R, Buchert R, Heinze H-J, Grace AA, Dolan RJ, Heinz A, Schlagenhauf F: **Ventral striatal dopamine reflects behavioral and neural signatures of model-based control during sequential decision making**. *Proc Natl Acad Sci* 2015, **112**:1595-1600.

49. Jarbo K, Verstynen TD: **Converging structural and functional connectivity of orbitofrontal, dorsolateral prefrontal, and posterior parietal cortex in the human striatum**. *J Neurosci* 2015, **35**:3865-3878.

50. Lee SW, Shimojo S, O'Doherty JP: **Neural computations underlying arbitration between model-based and model-free learning**. *Neuron* 2014, **81**:687-699.

51. Crockett MJ, Siegel JZ, Kurth-Nelson Z, Dayan P, Dolan RJ: **Moral
•• transgressions corrupt neural representations of value**. *Nat Neurosci* 2017, **20**:879.
This key study highlights a role for LPFC in the devaluation of money gained by harming others and provides a computational account of how individual differences in harm aversion can predict moral behavior.

52. Crockett MJ, Kurth-Nelson Z, Siegel JZ, Dayan P, Dolan RJ: **Harm to others outweighs harm to self in moral decision making**. *Proc Natl Acad Sci* 2014, **111**:17320-17325.

53. Crockett MJ, Siegel JZ, Kurth-Nelson Z, Ousdal OT, Story G, Frieband C, Grosse-Rueskamp JM, Dayan P, Dolan RJ: **Dissociable effects of serotonin and dopamine on the valuation of harm in moral decision making**. *Curr Biol* 2015, **25**:1852-1859.

54. Carlson RW, Aknin LB, Liotti M: **When is giving an impulse? An ERP investigation of intuitive prosocial behavior**. *Soc Cogn Affect. Neurosci* 2015, **11**:1121-1129.

55. Van Bavel JJ, FeldmanHall O, Mende-Siedlecki P: **The neuroscience of moral cognition: from dual processes to dynamic systems**. *Curr Opin Psychol* 2015, **6**:167-172.

56. Weisz E, Zaki J: **Motivated empathy: a social neuroscience perspective**. *Curr Opin Psychol* 2018, **24**:67-71.

57. Carlson RW, Marechal M, Oud B, Fehr E, Crockett M: **Motivated misremembering: selfish decisions are more generous in hindsight**. *PsyArXiv* 2018 http://dx.doi.org/10.31234/osf.io/7ck25. preprints.

58. Zaki J, Mitchell JP: **Intuitive prosociality**. *Curr Dir Psychol Sci* 2013, **22**:466-470.

59. Boyd R, Richerson PJ: **Punishment allows the evolution of cooperation (or anything else) in sizable groups**. *Ethol Sociobiol* 1992, **13**:171-195.

60. DeScioli P, Kurzban R: **Mysteries of morality**. *Cognition* 2009, **112**:281-299.

61. Baumgartner T, Knoch D, Hotz P, Eisenegger C, Fehr E: **Dorsolateral and ventromedial prefrontal cortex orchestrate normative choice**. *Nat Neurosci* 2011, **14**:1468.

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

63. Buckholtz JW, Asplund CL, Dux PE, Zald DH, Gore JC, Jones OD, Marois R: **The neural correlates of third-party punishment**. *Neuron* 2008, **60**:930-940.

64. Liu X, Li J, Wang G, Yin X, Li S, Fu X: **Transcranial direct current stimulation of the rLPFC shifts normative judgments in voluntary cooperation**. *Neurosci Lett* 2017. in press.

65. Buckholtz JW, Martin JW, Treadway MT, Jan K, Zald DH, Jones O,
•• Marois R: **From blame to punishment: disrupting prefrontal cortex activity reveals norm enforcement mechanisms**. *Neuron* 2015, **87**:1369-1380.
This research leverages NIBS to assess normative behavior and provides possibly the most compelling evidence to date for a causal role of DLPFC in norm enforcement behavior.

66. Eisenegger C, Treyer V, Fehr E, Knoch D: **Time-course of "off-line" prefrontal rTMS effects—a PET study**. *Neuroimage* 2008, **42**:379-384.

67. Swann NC, Tandon N, Pieters TA, Aron AR: **Intracranial electroencephalography reveals different temporal profiles for dorsal- and ventro-lateral prefrontal cortex in preparing to stop action**. *Cereb Cortex* 2013, **23**:2479-2488.

68. Batson CD, Ahmad N, Yin J, Bedell SJ, Johnson JW, Templin CM: **Two threats to the common good: self-interested egoism and empathy-induced altruism**. *Pers Soc Psychol Bull.* 1999, **25**:3-16.

69. 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.

70. Kahane G, Everett JAC, Earp BD, Caviola L, Faber NS, Crockett MJ, Savulescu J: **Beyond sacrificial harm: a two-dimensional model of utilitarian psychology**. *Psychol Rev* 2018, **125**:131-164.

71. Hutcherson CA, Montaser-Kouhsari L, Woodward J, Rangel A: **Emotional and utilitarian appraisals of moral dilemmas are encoded in separate areas and integrated in ventromedial prefrontal cortex**. *J Neurosci* 2015, **35**:12593-12605.

72. Monin B, Pizarro DA, Beer JS: **Reason and emotion in moral judgment: different prototypes lead to different theories**. *Do Emotions Help or Hurt Decision Making? A Hedgefoxian Perspective*. 2007.

73. Hofmann W, Meindl P, Mooijman M, Graham J: **Morality and self-control: how they are intertwined and where they differ**. *Curr Dir Psychol Sci.* 2018, **27**:286-291.

74. Badre D: **Cognitive control, hierarchy, and the rostro–caudal organization of the frontal lobes**. *Trends Cogn Sci* 2008, **12**:193-200.

75. Zinchenko O, Arsalidou M: **Brain responses to social norms: meta-analyses of fMRI studies**. *Human Brain Mapping* 2018, **39**:955-970.