Neural mechanisms of deliberate dishonesty: Dissociating deliberation from other control processes during dishonest behaviors

Liyang Sai[a,1,2], Gabriele Bellucci[b,1], Chongxiang Wang[a], Genyue Fu[a], Julia A. Camilleri[c,d], Simon B. Eickhoff[c,d], and Frank Krueger[e,2]

[a]Department of Psychology, Hangzhou Normal University, Hangzhou 311121, China; [b]Department of Computational Neuroscience, Max Planck Institute for Biological Cybernetics, Tübingen 72076, Germany; [c]Research Centre Jülich, Institute of Neuroscience and Medicine 52425 Jülich, Germany; [d]Institute for Systems Neuroscience, Medical Faculty, Heinrich Heine University, Düsseldorf 40225, Germany; and [e]School of Systems Biology, George Mason University, Fairfax, VA 22032

Edited by Thomas D. Albright, Salk Institute for Biological Studies, La Jolla, CA, and approved September 21, 2021 (received for review June 9, 2021)

Numerous studies have sought proof of whether people are genuinely honest by testing whether cognitive control mechanisms are recruited during honest and dishonest behaviors. The underlying assumption is: Deliberate behaviors require cognitive control to inhibit intuitive responses. However, cognitive control during honest and dishonest behaviors can be required for other reasons than deliberation. Across 58 neuroimaging studies (1,211 subjects), we investigated different forms of honest and dishonest behaviors and demonstrated that many brain regions previously implicated in dishonesty may reflect more general cognitive mechanisms. We argue that the motivational/volitional dimension is central to deliberation and provide evidence that motivated dishonest behaviors recruit the perigenual anterior cingulate cortex. This work questions the view that cognitive control is a hallmark of dishonesty.

dishonest | honesty | fMRI | metaanalysis | cognitive control

Different disciplines from philosophy, psychology, to neuroscience have tried to tackle the question as to whether individuals are intrinsically honest or dishonest. The central tenet underlying these efforts is that intuitive behavior reflecting one's "true" nature does not require cognitive control. Hence, if people are genuinely selfish and dishonest, deliberate cognitive control is needed to be honest (1, 2). On the contrary, if they are genuinely honest, cognitive control is required to behave dishonestly (3).

Methods such as time pressure or cognitive depletion are considered a good test bed for the above-mentioned hypotheses (3, 4). Intuitive behavior is thought to emerge when people have limited cognitive resources at their disposal to monitor, evaluate, and eventually change their behaviors. Neuroimaging studies have complemented these efforts by investigating whether people rely on brain areas associated with cognitive control during honest and dishonest behaviors, since some cognitive control processes are indicative of higher cognitive demands. However, results remain inconclusive. Some neuroimaging studies employing decision-making paradigms have observed neural activations in prefrontal cortical areas when people make honest decisions, like the anterior cingulate cortex (ACC), dorsolateral prefrontal cortex (DLPFC), and ventrolateral PFC (VLPFC) (5, 6)—regions associated with conflict monitoring, cognitive control, and response inhibition (7). Other studies, in which people were asked to lie about autobiographical and factual knowledge (6), have found similar neural activations but during dishonesty.

Even though identifying cognitive control areas as evidence of deliberation is a prima facie reasonable research objective, the implied assumption is questionable, as it implies that intuitive cheaters do not need to recruit cognitive control processes during dishonesty—a not entirely reasonable assumption. For instance, when reacting with a deceptive response, people still need to engage in counterfactual thinking to inhibit the truth and create alternative scenarios (8). Similarly, when responding honestly, intuitively honest individuals still need to engage in online monitoring and metacognitive processes due to self-image and reputational concerns (9).

Hence, it is unclear whether activations of brain regions associated with cognitive control in the literature are due to deliberation or other control processes. Cognitive demands unrelated to deliberation are present both when people are explicitly instructed to behave dishonestly (instructed dishonesty [ID]) and when they voluntarily engage in dishonest behaviors (spontaneous dishonesty [SD]). However, ID differs from SD because it does not involve the internal conflict inherent in a voluntary choice to be dishonest. Cognitive control mechanisms associated with deliberate dishonesty should hence be recruited only during SD. Finally, cognitive processes evoked by task-specific demands but unrelated to dishonesty should be found only in ID.

Results

We tested these hypotheses by identifying neuroimaging experiments that investigated honesty and dishonesty in paradigms that instructed participants how to behave (ID) and paradigms in which participants chose to act as they pleased (SD). First, the neural patterns consistently activated for dishonesty across all paradigms and ID were very similar, involving regions like the VLPFC, ACC, DLPFC, and inferior parietal lobule (IPL) (Fig. 1 *A* and *B*). On the contrary, honesty was found to consistently activate only the IPL.

Importantly, some of those activations might be related to other cognitive processes than deliberation, like the demands of following instructions. To single out neural activations closely related to deliberate dishonesty, we explored the differences between the neural patterns of ID and SD. Results reveal that ID more strongly recruits cognitive control brain areas than SD. Consensus connectivity maps indicate a partial overlap of the functional connectivity profile of these brain regions in the VLPFC (Fig. 2*D*). Hence, common cognitive control

Author contributions: L.S., G.B., and F.K. designed research; S.B.E. contributed new reagents/analytic tools; C.W. and J.A.C. analyzed data; L.S., G.B., and F.K. wrote the paper; and C.W., G.F., J.A.C., and S.B.E. reviewed and edited the paper.

The authors declare no competing interest.

[1]L.S. and G.B. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: liyangsai@hznu.edu.cn or fkrueger@gmu.edu.

This article contains supporting information online at http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2109208118/-/DCSupplemental.

Published October 18, 2021.

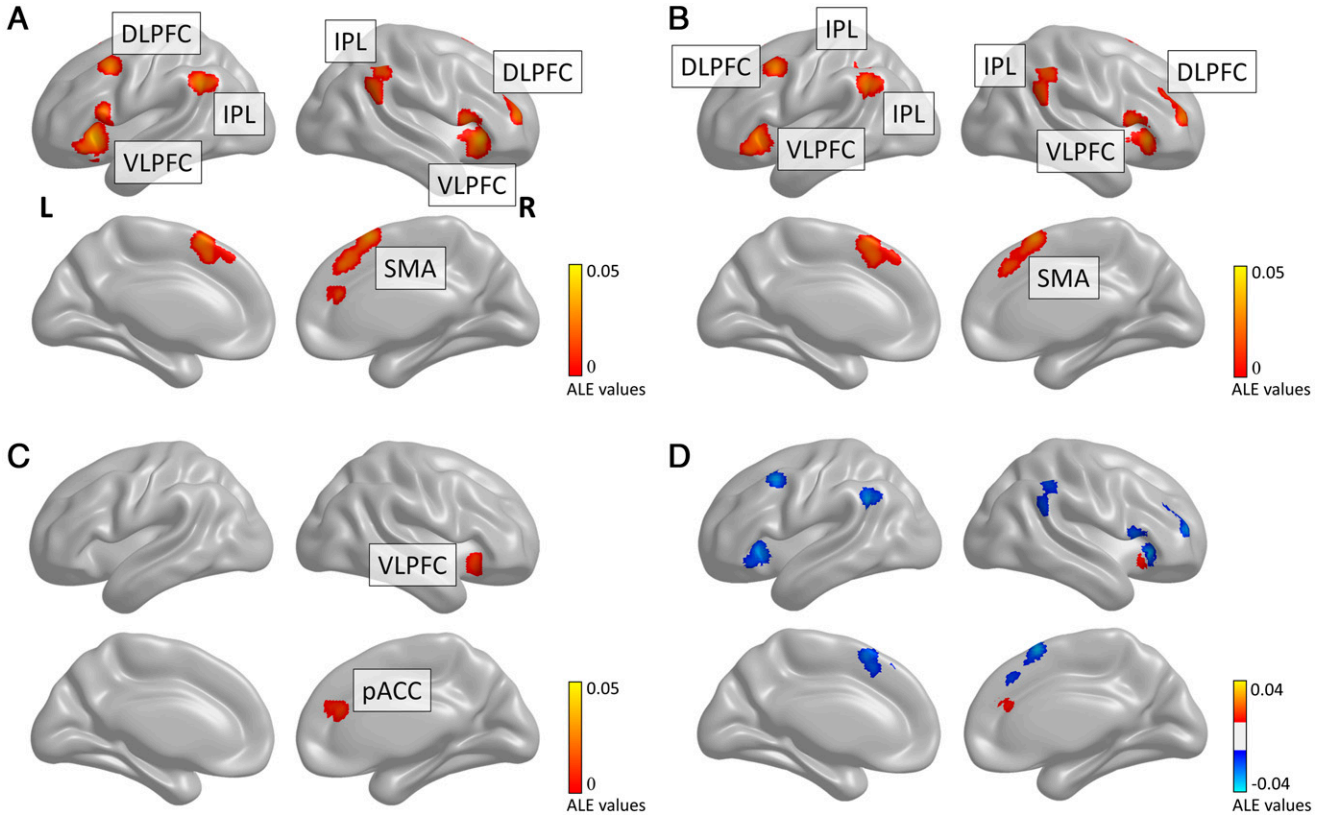

Fig. 1. Brain regions underlying dishonesty. Recruitment of cognitive control brain areas for dishonesty (cluster family-wise error [cFWE] < 0.05) (*A*). Main effects of instructed (*B*) and spontaneous dishonesty (*C*), and their contrast (*D*). L, left; R, right; IPL, inferior parietal lobule; SMA, supplementary motor area.

mechanisms likely related to the shared demands of dishonesty are evoked by both ID and SD with recruitment of the VLPFC. However, ID requires additional control processes, likely related to other task demands.

Clustering analyses provide insights into the neural patterns associated with these demands. Specifically, the connectivity profile of the brain regions activated by ID clusters into a specific subnetwork with a hub in the DLPFC (Fig. 2*B*). This

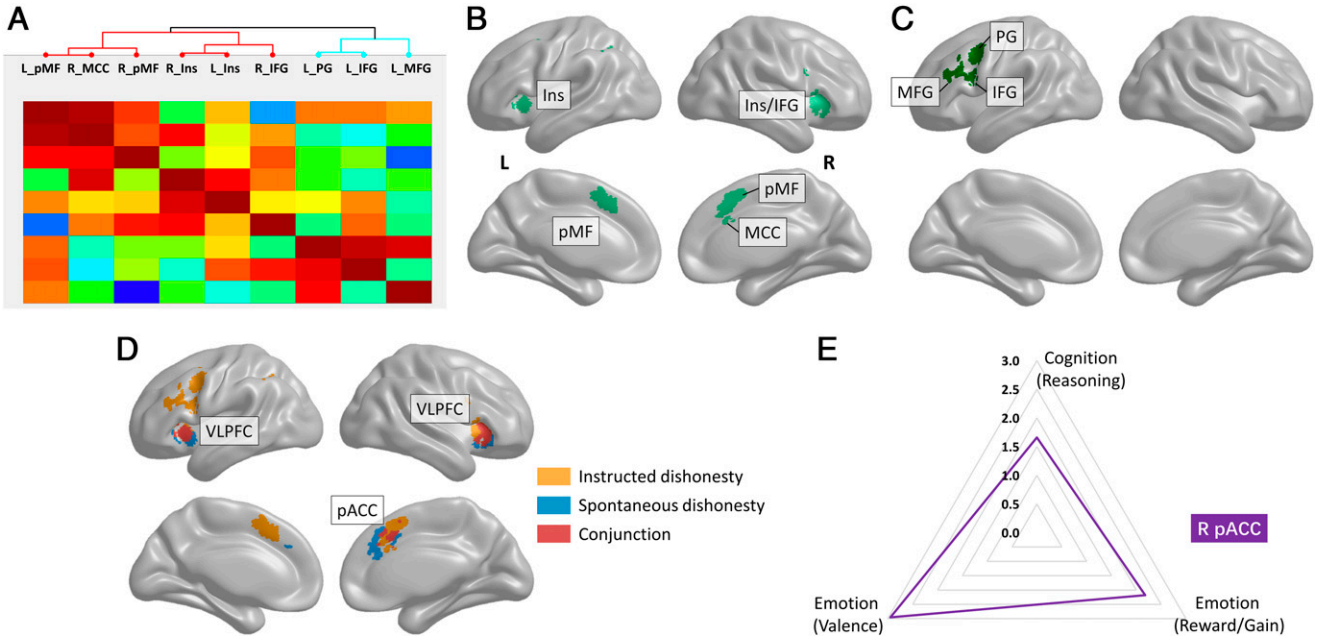

Fig. 2. ACC connectivity and functions during ID and SD. (*A*) Hierarchical clustering analysis with subnetworks (red/blue). (*B* and *C*) ID subnetworks (light/dark green). (*D*) Consensus connectivity analyses for ID (orange), SD (blue), and their conjunction (red). (*E*) Functional decoding analyses (subcategories in parentheses). L, left; R, right; PG, precentral gyrus; IFG, inferior gyrus; MFG, middle frontal gyrus; Ins, insula; pMF, posterior-medial frontal gyrus; MCC, middle cingulate cortex.

DLPFC connectivity network was found only during ID and was distinct from a second subnetwork more similar to SD's neural patterns (Fig. 2*C*). This suggests that brain regions (e.g., DLPFC) that are classically linked to deliberate processes during dishonest behaviors are not related to deliberate dishonesty.

As processes associated with a deliberate dishonest decision are inherent to situations implying an internal conflict, they are absent (or mildly present) when individuals are instructed on how to behave. On the contrary, they arise when people voluntarily choose what to do (i.e., in spontaneous dishonest behaviors). Analyses of SD neural patterns show consistent activations only in the perigenual ACC (pACC) and VLPFC (Fig. 1 *C* and *D*, *SI Appendix*). Functional decoding analyses revealed that pACC activations were associated with negatively valenced emotions and cognitive functioning, in line with an internal conflict during motivated dishonesty (Fig. 2*E*).

Discussion

By investigating contexts where people behave as they please (spontaneous behaviors), we identified the distinctive, volitional dimension of deliberation—largely overlooked by previous work due to the focus on other cognitive control processes (down-regulation/suppression of intuitive responses). For SD, instead of regions classically associated with regulatory or inhibitory mechanisms (e.g., DLPFC), we observed consistent activations in the pACC, a region more closely related to volition and motivation (10). These findings indicate that identifying cognitive control brain regions is not sufficient proof for deliberate behavior, which requires contexts allowing people to voluntarily choose how to behave. This aligns with the view of dishonest behaviors in other domains like the legal domain, where prosecution requires assessing whether the perpetrator has a mens rea (guilty intent) (11).

An open question is whether our results were driven by an overwhelming presence of honest subjects in our dataset (5, 12) and would hold in less homogenous populations. For instance, a recent neuroimaging study suggests cognitive control is needed to cheat for honest people, but to be honest for cheaters (12). However, our results are consistent with this study, which lacks a proper control of cognitive control mechanisms unrelated to deliberation. In particular, since the brain adapts to dishonesty (13), honest individuals who cheat rarely might require more cognitive control to switch from honest to dishonest behavior than frequent cheaters for whom continuing to cheat might be less cognitively effortful.

Moreover, considerations about the social consequences of one's dishonest behavior, particularly whether someone else is hurt by it, are central to a deliberate choice of being dishonest (14). However, since previous neuroscientific and behavioral studies employ largely nonsocial paradigms, this social dimension has been chiefly neglected and needs to be considered in future research.

Finally, the relatively low number of studies for SD compared to ID might have prevented us from detecting effects of small effect size (15). Hence, as absence of evidence does not imply evidence of absence, we cannot exclude that other brain regions, specifically those traditionally associated with cognitive control, might still have some small contribution to SD. A priority in future research is to directly test the involvement of brain regions traditionally associated with cognitive control by running single experiments and additional metaanalyses with more studies.

Materials and Methods

First, relevant articles were found after a literature review according to Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. Next, activation likelihood estimation metaanalyses were performed for honesty/dishonesty. Finally, metaanalytical connectivity modeling and resting-state functional connectivity analyses were used to examine the coactivation and functional connectivity patterns of the metaanalytic results, respectively. Functional decoding analyses were used for functional profile characterization. Details are provided in the *SI Appendix*.

Data Availability. Data have been deposited in the Open Science Framework (https://osf.io/dne3r/) (16) and all other data are included in *SI Appendix*.

ACKNOWLEDGMENTS. We thank Tingwen Sheng for helping to screen the literature.

1. J. Haidt, The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychol. Rev.* **108**, 814–834 (2001).
2. D. Kahneman, *Thinking, Fast and Slow* (Farrar, Straus and Giroux, ed. 1, 2013).
3. B. Verschuere, N. C. Köbis, Y. Bereby-Meyer, D. Rand, S. Shalvi, Taxing the brain to uncover lying? Meta-analyzing the effect of imposing cognitive load on the reaction-time costs of lying. *J. Appl. Res. Mem. Cogn.* **7**, 462–469 (2018).
4. S. Shalvi, O. Eldar, Y. Bereby-Meyer, Honesty requires time (and lack of justifications). *Psychol. Sci.* **23**, 1264–1270 (2012).
5. J. D. Greene, J. M. Paxton, Patterns of neural activity associated with honest and dishonest moral decisions. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 12506–12511 (2009).
6. N. Lisofsky, P. Kazzer, H. R. Heekeren, K. Prehn, Investigating socio-cognitive processes in deception: A quantitative meta-analysis of neuroimaging studies. *Neuropsychologia* **61**, 113–122 (2014).
7. H. Garavan, T. J. Ross, E. A. Stein, Right hemispheric dominance of inhibitory control: An event-related functional MRI study. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 8301–8306 (1999).
8. R. A. Briazu, C. R. Walsh, C. Deeprose, G. Ganis, Undoing the past in order to lie in the present: Counterfactual thinking and deceptive communication. *Cognition* **161**, 66–73 (2017).
9. S. M. Rosenbaum, S. Billinger, N. Stieglitz, Let's be honest: A review of experimental evidence of honesty and truth-telling. *J. Econ. Psychol.* **45**, 181–196 (2014).
10. G. Bellucci, J. A. Camilleri, S. B. Eickhoff, F. Krueger, Neural signatures of prosocial behaviors. *Neurosci. Biobehav. Rev.* **118**, 186–195 (2020).
11. F. Krueger, M. Hoffman, The emerging neuroscience of third-party punishment. *Trends Neurosci.* **39**, 499–501 (2016).
12. S. P. H. Speer, A. Smidts, M. A. S. Boksem, Cognitive control increases honesty in cheaters but cheating in those who are honest. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 19080–19091 (2020).
13. N. Garrett, S. C. Lazzaro, D. Ariely, T. Sharot, The brain adapts to dishonesty. *Nat. Neurosci.* **19**, 1727–1732 (2016).
14. N. C. Köbis, B. Verschuere, Y. Bereby-Meyer, D. Rand, S. Shalvi, Intuitive honesty versus dishonesty: Meta-analytic evidence. *Perspect. Psychol. Sci.* **14**, 778–796 (2019).
15. S. B. Eickhoff *et al.*, Behavior, sensitivity, and power of activation likelihood estimation characterized by massive empirical simulation. *Neuroimage* **137**, 70–85 (2016).
16. L. Sai *et al.*, Neural mechanisms of deliberate dishonesty: Dissociating deliberation from other control processes during dishonest behaviors. Open Science Framework. https://osf.io/dne3r (Accessed 7 October 2021).