Neuroscience of Moral Decision Making

Yang Hu[a,b,*], Xiaoxue Gao[a,b,*], Hongbo Yu[c], Zhewen He[b,d], and Xiaolin Zhou[a,b,e], ,
r r , , ' u , ,
r , , ' u , D r , r r , r r , A,
, D , r , , , -D / r u r r
r , r , , ' u

Introduction

As moral agents, we human beings are equipped with the capability to make judgments about the moral appropriateness of the other's behaviors (i.e., moral judgment) (Baron, 2014; Malle et al., 2014; Wojciszke et al., 2015
very often raised by philosophers, social scientists (especially psychologists) and the general public. Rather than coming
mprehensive and commonly agreed definition of morality (or moral domain), researchers often define morality from
es (Bartels et al., 2014; Crockett, 2013; Haidt, 2007). Two of those are commonly adopted. The first approach highlights
ce with norms, which describe expectancies, beliefs, and rules (Graham et al., 2011 2013). The other addresses the concern for the other's welfare, typically accompanied
ieving the other's suffering) (Cushman, 2015). Accordingly, immorality refers to the behaviors

2012) and the underlying neural activation patterns (FeldmanHall et al., 2012; Gospic et al., 2013) are different in hypothetical versus real contexts. Moreover, most of these previous studies were not designed to provide a mechanistic account for the moral behaviors,

the partner), which was measured by the no-provocation trials in the Taylor reaction-time aggression paradigm (Taylor, 1967).

Notably, all studies mentioned above assumed that helping behaviors would surely reduce the others' suffering, which was not always true in real life. To address this issue, a study combining both fMRI and tDCS techniques developed a new paradigm in which

implying that higher demands in moral mentalizing are required in social decision-making when the decision to reject could not be readily justi

revealing that the resting-state brain activity in the left ventral AI (as well as other regions) was correlated with the PIF response (Xu et al., 2015). Together, these findings suggest that the AI is not only engaged in signaling social norm violation during UG but also recruited in guiding subsequent adaptive behaviors (e.g., PIF response).

Learning

In real life, we not only make moral choices in one shot, but often need to form and update our beliefs about the moral trait of others, thereby guiding how we should get along with them in the future (Siegel et al., 2018). Although a substantial amount of evidence has revealed the neurocomputational mechanisms underlying how people learn through feedbacks under the general framework of reinforcement learning (O'Doherty et al., 2017), the neural underpinnings through which we infer the moral character of other people are still poorly understood. To investigate this issue, Hackel et al. (2015) performed a fMRI study in which participants were asked to learn how generous an anonymous partner was via trial-and-error learning based on the proportion of resources shared by the partner. As a control condition, participants also needed to learn which slot machine earned themselves more. Model-based analyses revealed that participants relied more on generosity information than on reward value during the task. Trial-wise prediction error (PE) of both types of information was commonly encoded in the right VS. However, the generosity prediction error recruited an additional network in association with the formation of social impression, including the ventral lateral prefrontal cortex (vlPFC), IPL, PCC extending to precuneus, as well as the right TPJ. Another study with a similar learning paradigm also found a signal of generosity PE in the PCC/precuneus (Stanley, 2016). Furthermore, our ability to infer others' moral character

The third issue is related to methodological approaches that should be taken to provide additional information from different viewpoints, thereby characterizing a panoramic view of the moral brain. Obviously, the current literature predominantly considers which parts of the brain (and the inter-regional connections) are associated with a specific form of moral decision using fMRI, supplemented by the causality methods such as brain lesion and non-invasive brain stimulation (e.g., TMS, tDCS). There have been several studies adopting the EEG technique (e.g., event-related potential, ERP) to explore the temporal features of moral decision

Garrett, N., Lazzaro, S.C., Ariely, D., Sharot, T., 2016. The brain adapts to dishonesty. Nat. Neurosci. 19 (12), 1727.
Garrigan, B., Adlam, A.L., Langdon, P.E., 2016. The neural correlates of moral decision-making: a systematic review and meta-analysis of moral evaluations and response decision judgements. Brain Cognit. 108, 88–97.
Gert, B., 2004. Common Morality: Deciding what to Do. Oxford University Press.
Ginther, M.R., Bonnie, R.J., Hoffman, M.B., Shen, F.X., Simons, K.W., Jones, O.D., Marois, R., 2016. Parsing the behavioral and brain mechanisms of third-party punishment. J. Neurosci. 36 (36), 9420–9434.
Gneezy, U., 2005. Deception: the role of consequences. Am. Econ. Rev. 95 (1), 384–394.
Gospic, K., Sundberg, M., Maeder, J., Fransson, P., Petrovic, P., Isacsson, G., et al., 2013. Altruistic motives from amygdala. Soc. Cognit. Affect Neurosci. 9 (9), 1325–1332.
Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S.P., Ditto, P.H., 2013. Moral foundations theory: the pragmatic validity of moral pluralism. Adv. Exp. Soc. Psychol. 47, 55–130.
Graham, J., Nosek, B.A., Haidt, J., Iyer, R., Koleva, S., Ditto, P.H., 2011. Mapping the moral domain. J. Pers. Soc. Psychol. 101 (2), 366.
Grecucci, A., Giorgetta, C., Van't Wout, M., Bonini, N., Sanfey, A.G., 2012. Reappraising the ultimatum: an fMRI study of emotion regulation and decision making. Cerebr. Cortex 23 (2), 399–410.
Greene, J.D., 2015. The cognitive neuroscience of moral judgment and decision-making. In: Gazzaniga, M.S., Wheatley, T. (Eds.), The Cognitive Neurosciences, A multidisciplinary perspective, pp. 197–220. Boston (Review).
Greene, J.D., Paxton, J.M., 2009. Patterns of neural activity associated with honest and dishonest moral decisions. Proc. Natl. Acad. Sci. U. S. A. 106 (30), 12506–12511.
Guo, X., Zheng, L., Zhu, L., Li, J., Wang, Q., Dienes, Z., Yang, Z., 2013. Increased neural responses to unfairness in a loss context. Neuroimage 77, 246–258.
Gürglu, B., van den Bos, W., Rombouts, S.A., Crone, E.A., 2010. Unfair? It depends: neural correlates of fairness in social context. Soc. Cognit. Affect Neurosci. 5 (4), 414–423.
Gürglu, B., Will, G.-J., Crone, E.A., 2014. Neural correlates of advantageous and disadvantageous inequity in sharing decisions. PLoS One 9 (9), e107996.
Haber, S.N., Knutson, B., 2009. The reward circuit: linking primate anatomy and human imaging. Neuropsychopharmacology 35 (1), 4–26.
Hackel, L.M., Amodio, D.M., 2018. Computational neuroscience approaches to social cognition. Curr. Opin. Psychol. 24, 92–97.
Hackel, L.M., Doll, B.B., Amodio, D.M., 2015. Instrumental learning of traits versus rewards: dissociable neural correlates and effects on choice. Nat. Neurosci. 18 (9), 1233–1235.
Haidt, J., 2003. The moral emotions. In: Schrerer, K.R., Goldsmith, H.H. (Eds.), Handbook of Affective Sciences, vol. 11. Oxford University Press, Oxford, pp. 852–870.
Haidt, J., 2007. The new synthesis in moral psychology. Science 316 (5827), 998–1002.
Haidt, J., 2008. Morality. Perspect. Psychol. Sci. 3 (1), 65–72.
Harbaugh, W.T., Mayr, U., Burghart, D.R., 2007. Neural responses to taxation and voluntary giving reveal motives for charitable donations. Science 316 (5831), 1622–1625.
Hare, T.A., Camerer, C.F., Knoepfle, D.T., O'Doherty, J.P., Rangel, A., 2010. Value computations in ventral medial prefrontal cortex during charitable decision making incorporate input from regions involved in social cognition. J. Neurosci. 30 (2), 583–590.
Hare, T.A., Schultz, W., Camerer, C.F., O'Doherty, J.P., Rangel, A., 2011. Transformation of stimulus value signals into motor commands during simple choice. Proc. Natl. Acad. Sci. U. S. A. 108 (44), 18120–18125.
Haruno, M., Frith, C.D., 2010. Activity in the amygdala elicited by unfair divisions predicts social value orientation. Nat. Neurosci. 13 (2), 160–161.
Hauser, M., Lee, J., Huebner, B., 2010. The moral-conventional distinction in mature moral competence. J. Cognit. Cult. 10 (1–2), 1–26.
Hsu, M., Anen, C., Quartz, S.R., 2008. The right and the good: distributive justice and neural encoding of equity and efficiency. Science 320 (5879), 1092–1095.
Hu, J., Hu, Y., Li, Y., Zhou, X., 2021. Computational and neurobiological substrates of strategic helping decision. J. Neurosci. 41 (15), 3535–3545.
Hu, J., Li, Y., Yin, Y., Blue, P.R., Yu, H., Zhou, X., 2018. How do self-interest and other-need interact in the brain to determine altruistic behavior? Neuroimage 167, 197–208.
Hu, Y., He, L., Zhang, L., Wolk, T., Dreher, J.C., Weber, B., 2018. Spreading inequality: neural computations underlying paying it forward altruism. Soc. Cognit. Affect. Neurosci. 13 (6), 575–589. https://doi.org/10.1093/scan/nsy040
Hu, Y., Pereira, A.M., Gao, X., Campos, B.M., Derrington, E., Corgnet, B., et al., 2021. Right temporoparietal junction underlies avoidance of moral transgression in autism spectrum disorder. J. Neurosci. 41 (8), 1699–1719. https://doi.org/10.1523/JNEUROSCI.1237-20.2020
Hu, Y., Scheele, D., Becker, B., Voos, G., David, B., Hurlemann, R., Weber, B., 2016. The effect of oxytocin on third-party altruistic decisions: an fMRI Study. Sci. Rep. 6, 20236. https://doi.org/10.1038/srep20236
Hu, Y., Strang, S., Weber, B., 2015. Helping or punishing strangers: neural correlates of altruistic decisions as third-party and its relation to empathic concern. Front. Behav. Neurosci. 9, 24. https://doi.org/10.3389/fnbeh.2015.00024
Hutcherson, C., Bushong, B., Rangel, A., 2015. A neurocomputational model of altruistic choice and its implications. Neuron 87 (2), 451–462.
Hwang, C.-L., Lin, M.-J., 2012. Group Decision Making under Multiple Criteria: Methods and Applications, vol. 281. Springer Science & Business Medi
Ismayilov, H., Potters, J.J.J., 2015. Promises as Commitments.
Izuma, K., Saito, D.N., Sadato, N., 2010. Processing of the incentive for social approval in the ventral striatum during charitable donation. J. Cognit. Neurosci. 22 (4), 621–631.
Kahneman, D., Knetsch, J.L., Thaler, R., 1986a. Fairness as a constraint on profit seeking: entitlements in the market. Am. Econ. Rev. 728–741.
Kahneman, D., Knetsch, J.L., Thaler, R.H., 1986b. Fairness and the assumptions of economics. J. Bus. S285–S300. https://doi.org/10.1086/296367

Ma, Q., Hu, Y., Jiang, S., Meng, L., 2015. The undermining effect of facial attractiveness on brain responses to fairness in the Ultimatum Game: an ERP s
9, 77.
Malle, B.F., Guglielmo, S., Monroe, A.E., 2014. A theory of blame. Psychol. Inq. 25 (2), 147
Maréchal, M.A., Cohn, A., Ugazio, G., Ruff, C.C., 2017. Increasing honesty in humans with noninvasive brain stimulation. Proc. Natl. Acad. Sci. U. S.
Masserman, J.H., Wechkin, S., Terris, W., 1964. "Altruistic" behavior in rhesus monkeys. Am. J. Psychiatr. 121 (6), 584
McAuliffe, K., Blake, P.R., Steinbeis, N., Warneken, F., 2017. The developmental foundations of human fairness. Nat. Human Behav. 1 (2), 0042.
McCullough, M.E., Kilpatrick, S.D., Emmons, R.A., Larson, D.B., 2001. Is gratitude a moral affect? Psychol. Bull. 127 (2), 249.
Moll, J., De Oliveira-Souza, R., Zahn, R., 2008. The neural basis of moral cognition: sentiments, concepts, and values. Ann. N. Y. Acad. Sci. 1124 (1),
Moll, J., Krueger, F., Zahn, R., Pardini, M., de Oliveira-Souza, R., Grafman, J., 2006. Human fronto-mesolimbic networks guide decisions about charitable donation. Proc. Natl.

van Baar, J.M., Chang, L.J., Sanfey, A.G., 2019. The computational and neural substrates of moral strategies in social decision-making. Nat. Commun 10 (1), 1483.
van Den Bos, W., van Dijk, E., Westenberg, M., Rombouts, S.A., Crone, E.A., 2009. What motivates repayment? Neural correlates of reciprocity in the Trust Game. Soc. Cogn. Affect Neurosci. 4 (3), 294–304.
van den Bos, W., van Dijk, E., Westenberg, M., Rombouts, S.A., Crone, E.A., 2011. Changing brains, changing perspectives: the neurocognitive development of reciprocity. Psychol. Sci. 22 (1), 60–70.
Volz, K.G., Vogeley, K., Tittgemeyer, M., von Cramon, D.Y., Sutter, M., 2015. The neural basis of deception in strategic interactions. Front. Behav. Neurosci. 9, 27.
Watanabe, T., Takezawa, M., Nakawake, Y., Kunimatsu, A., Yamasue, H., Nakamura, M., et al., 2014. Two distinct neural mechanisms underlying indirect reciprocity. Proc. Natl. Acad. Sci. U. S. A. 111 (11), 3990–3995.
Wojciszke, B., Parzuchowski, M., Bocian, K., 2015. Moral judgments and impressions. Curr. Opin. Psychol. 6, 50–54.
Woo, C.-W., Koban, L., Kross, E., Lindquist, M.A., Banich, M.T., Ruzic, L., Andrews-Hanna, J.R., Wager, T.D., 2014. Separate physical pain and social rejection. Nat. Communi. 5, 5380.
Wrangham, R.W., 2018. Two types of aggression in human evolution. Proc. Natl. Acad. Sci. U. S. A. 115 (2), 245–253.
Wu, Y., Leliveld, M.C., Zhou, X., 2011. Social distance modulates recipient