How lesioned studies and neuroimaging improve our understanding in social cognition

Lesioned Studies

Studying the effects of brain lesions in humans can reveal the building blocks of cognition and update the information processing framework. Decision-making processes differ in social and non-social contexts, with lesion studies shedding light on these processes.

Social decision-making differs from non-social decision-making in three processes (Lee & Harris, 2013), (1) making predictions guiding decision-making, (2) examining decision outcomes, and (3) using outcomes for learning. In the learning phase, individuals may form initial impressions, which affect behaviour in social contexts (van ’t Wout & Sanfey, 2008). Confirmatory bias, where people seek information consistent with pre-existing beliefs, may play a role in the economic games used in social decision-making studies, as partners often provide probabilistic (and sometimes ambiguous) feedback. The interpretation of this feedback may be influenced by prior beliefs (Delgado et al., 2005).

A cognitive psychology study using economic games and computational modelling found that initial impressions and prior interactions influence behaviour significantly (Chang et al., 2010). They used Reinforcement Learning (RL) models to test several distinct processing hypotheses and found that trustworthiness is believed to be a probability of reciprocation initially based on implicit judgments and dynamically updated based on experiences. However, whether the learning processes of human partners and machines share the same mechanism, differing only due to inputs, or whether these differences stem from distinct processes, is a question that cannot be answered using cognitive psychology methods. Nevertheless, lesion studies can provide insights into this matter.

Zhu et al. (2019) compared patients with lesions in the orbitofrontal cortex (OFC) or basal ganglia (BG) with healthy subjects. The study revealed that while all patients were impaired when playing against a computer, BG patients demonstrated preserved learning when playing against humans. This suggests the existence of higher-order learning processes supported by OFC, dissociable from trial-and-error learning.

In this study, subjects were instructed to participate in an economic game called Patent Race. They were endowed with either five or four units at the beginning of each round and were tasked to decide how many units from this newly endowed set to invest in order to compete with a rival. If their investment exceeded the rival's, they would win a prize (worth 10 units) along with the units left. Conversely, if their investment did not exceed the rival's, they would only retain the units left. In the strategic context condition, subjects played the game against a human, while in the non-strategic context condition, the rival was replaced by a computer algorithm. Subjects completed 80 rounds per condition, with each round randomly matched. This randomization ensured that, in the strategic context condition, subjects could not specifically learn a particular person's mental states, thereby minimizing the role of reputation and higher-order belief considerations.

To examine whether competing against a human or a computer algorithm elicited different learning processes, the authors calculated the stay rate. This rate measures the probability of subjects choosing to invest the same units in the next round, considering the corresponding outcome. Subjects could apply a win-stay-loss-switch strategy in both reinforce-based learning and belief-based learning. However, only the belief-based learning process was sensitive to whether the subject's investment exceeded their rival's investment, known as regret, which refers to the difference between the potential maximum payoff and the actual payoff (Camerer & Hua Ho, 1999).

The findings indicated that in the strategic context, both BG patients and healthy controls (HC) exhibited a higher stay rate when they earned a high payoff compared to a low payoff. Similarly, they showed a lower stay rate when they had high regret than low regret. However, OFC patients showed no significant differences. This suggests that HC and BG patients applied belief-based learning in the strategic context. In the non-strategic context, HC and OFC patients showed a higher stay rate when they earned a high payoff compared to a low payoff. Yet, BG patients did not exhibit any significant difference. None of the HC, OFC patients, or BG patients showed any significant differences between high and low regret (Fig. 1A). This suggests that HC and OFC patients applied reinforce-based learning in the non-strategic context.

Furthermore, the authors applied a computational modelling method. They employed a well-established model, the experience-weighted attraction (EWA) model (Camerer & Hua Ho, 1999), which describes a hybrid of reinforce-based learning and belief-based learning, and a basic RL model to fit their data.

Their findings revealed that in the strategic context, the pseudo-R2, a measure of model fit of the EWA model was significantly higher in HC and BG patients than in OFC patients. However, in the non-strategic context, the pseudo-R2 of the EWA model was significantly higher in HC than in BG and OFC patients (Fig. 1B). Utilizing the Bayesian Information Criterion (BIC) to compare the EWA and RL models, they discovered that in the strategic context, the EWA model outperformed the RL model significantly in HC and BG patients but not in OFC patients. Conversely, in the non-strategic context, the RL model performed significantly better than the EWA model in OFC patients but not in HC and BG patients (Fig. 1C).

In summary, they observed a dissociation between reinforce-based learning and belief-based learning. This dissociation cannot be explained by a higher requirement of neural resources for reinforce-based learning compared to belief-based learning. This is theoretically reversed, as belief-based learning is considered a higher-order process (Evdokimov & Garfagnini, 2022).

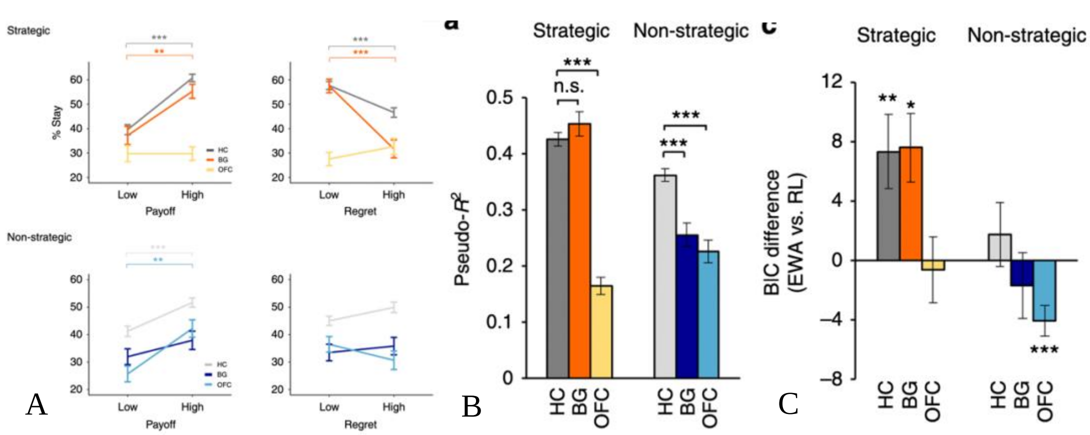

Fig 1. Result of Zhu et al. (2019).

While people already understand that the difference in the learning process during social and non-social decision-making can be described as reinforce-based learning and belief-based learning, it is challenging to determine whether they represent two distinct systems or two conditions of a single system. Specifically, the EWA model, a well-established framework that delineates reinforce-based learning and belief-based learning, treats pure reinforce-based learning or belief-based learning as specific scenarios (Camerer & Hua Ho, 1999). Some scholars have found that differentiating between various learning models based solely on behavioural choice data is exceedingly challenging (Salmon, 2001; Wilcox, 2006). However, lesion studies indicate that when a social context is present, allowing for the anticipation of future actions by others, belief-based learning can serve as a compensatory process. This unique insight suggests that the reinforce-based learning process and belief-based learning process function as isolated learning processes (Fig. 2).

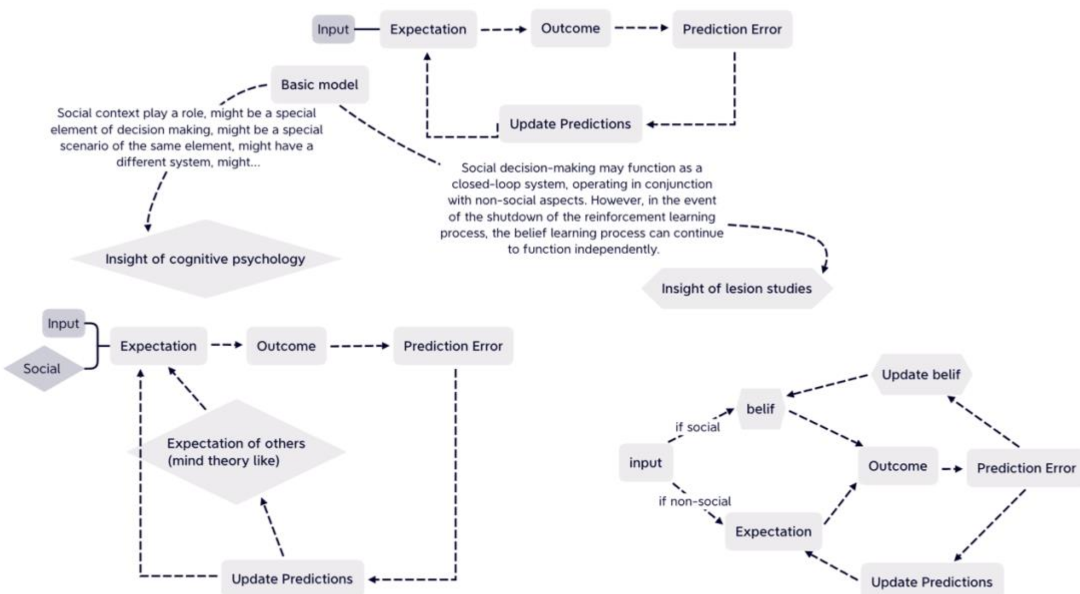

Figure 2. Different contribution made by cognitive psychology and lesion study in explaining why social decision making different.

Another instance where lesion studies enhance our comprehension of the social decision-making process involves the trade-off between honesty and dishonesty. Two longstanding hypotheses on dishonesty/deception are the Will hypothesis and the Grace hypothesis (Greene & Paxton, 2009; for review, see Speer et al., 2022). The former suggests that people's innate nature tends toward selfishness and dishonesty, requiring deliberate cognitive control to maintain honesty, akin to exercising self-control when delaying gratification. Conversely, the latter proposes that individuals are naturally honest, and occasionally, cognitive processes benefit from opportunities to deceive. Cognitive psychology studies can disentangle these two processes (Fosgaard et al., 2013) and have discovered their susceptibility to manipulation through cognitive loading. However, they cannot precisely discern how these cognitive processes affect decision-making. Does the Grace process amplify a general reward effect that could also be inhibited by the Will process (Fig. 3A)? Or do they specifically influence a cheating-related reward effect (Fig. 3B)? This question can be addressed through lesion studies.

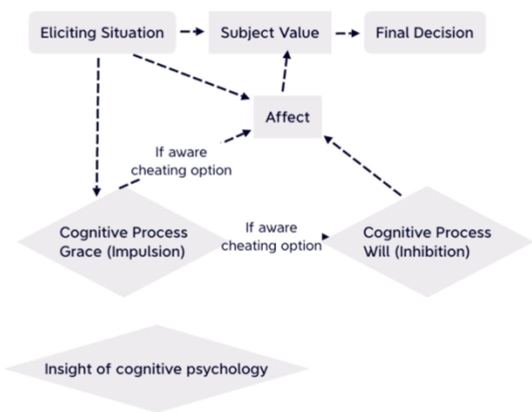

Figure 3. Different contribute made by cognitive psychology and lesion study in explaining how cognitive control affect cheating behaviour.

In another study conducted by Zhu et al. (2014) , they compared dorsolateral prefrontal cortex (dlPFC) patients, OFC patients, and HC. Patients were presented with two conditions: in the choice condition, they could select a payoff allocation plan without any opportunity for cheating, while in the message condition, they had the option to send a false message to a partner for additional money (Fig. 4a). The findings revealed that when subjects had no chance to cheat, they allocated a similar amount of money to their partners. However, when they could send a false message, dlPFC patients tended to give much less money (indicating more cheating) to their partners compared to OFC patients and HC (Fig. 4b). Moreover, when the interests of the subject and their partner did not conflict, dlPFC patients behaved as honestly as OFC patients and HC. Computational modelling demonstrated that while dlPFC patients valued their partners equally to others in the choice condition, in the message condition, those with fully functioning dlPFC exhibited a diminished inclination to use cheating as a means to earn as much money as they would in the choice condition (Fig. 4d).

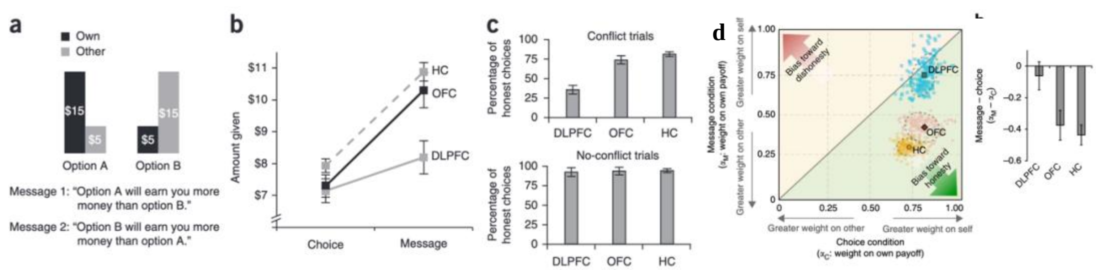

Fig 4. Result of Zhu et al. (2014).

Comparing lesion patients with other patients having lesions and healthy controls allows us to understand the potential outcomes when one or multiple cognitive elements are blocked. This provides insights into the sequence and structure of the cognitive processes involved in a particular task. Eventually, in combination with psychology and neuroimaging studies, lesion studies can help answer the question of how our minds work.

Neuroimaging

The most significant difference between neuroimaging and lesion studies lies in their differing inferential logic. Functional imaging studies instruct subjects to complete various tasks and then compare brain activities. By selecting an appropriate baseline task, researchers can infer which brain region may be sufficient for performing several tasks. In contrast, lesion studies compare patients and healthy subjects in the same tasks to infer which brain region is necessary for performing those tasks. Rather than simply switching the on-off state of certain brain regions, neuroimaging provides a unique opportunity to measure their volume. Consequently, we can use regression or multivoxel pattern analysis (MVPA) to gather more information about how brains process information.

Cognitive psychology studies have revealed numerous cognitive processes. However, they are limited in their ability to reveal nuances in automatic and implicit processes through subjective measures and semantic tasks. Functional imaging studies can offer objective measures of cognitive processes and opportunities to distinguish components that may be challenging to dissociate using psychological methods but easily observable in imaging data.

One example in social perception could be inequity aversion. A longstanding question in economic and social psychology is why people don't act as pro-self as rational choice theory suggests (Boudon, 2003). It is suggested that humans tend to reduce inequality in outcome distributions (Bolton & Ockenfels, 2000; Fehr & Schmidt, 1999). However, they still cannot answer the question of why humans dislike inequity. Do individuals care directly about inequality, or do they infer intentions from actions that cause unequal outcomes? Particularly, when subjects are better-off, their behaviours of avoiding inequality may be due to social image (Andreoni & Bernheim, 2009) or reciprocity (Rabin, 1993) instead of a pure aversion to inequality.

Through a refined design, psychology experiments (Dawes et al., 2007) can prevent the development of reputations, reinforce cooperation, and prevent retaliations, revealing pure inequality aversion. However, they cannot demonstrate whether advantageous and disadvantageous inequality are processed in the same way.

Using fMRI and economic games, Tricomi et al. (2010) found that the ventral striatum (VS) and ventromedial prefrontal cortex (vmPFC) activated more when achieving more equity by giving or taking. In this study, subjects were randomly matched to form a dyad. One in each pair was randomly chosen to gain an extra bonus (high-pay), whereas the other received no bonus (low-pay) (Fig. 5a). Then, during fMRI scanning, several proposals of random money offers were made to the low-pay subject, high-pay subject, or both. Subjects were instructed to rate how appealing those proposals were (Fig. 5b). By regressing appealing points on the offered money, we could discern the subject's value of transferring money from the experimenter to themselves or others. In this study, high-paid subjects showed a positive subjective value for transfer to others, whereas low-paid subjects showed a negative value, signifying inequity aversion (Fig. 5c).

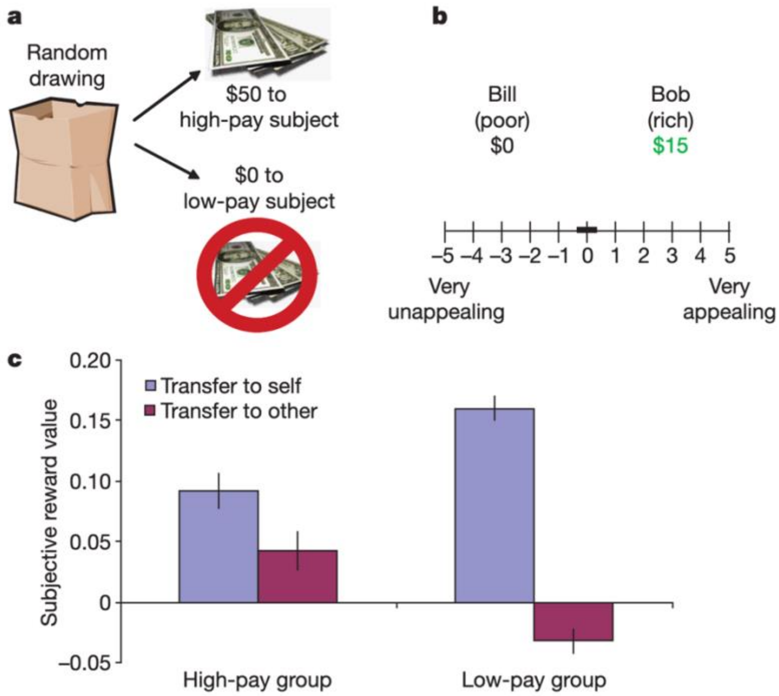

Fig. 5 Paradigm and behavioural result (Tricomi et al., 2010)

Neural data shows that the VS, linked to processing reward information (Schultz, 2002), was more activated in all subjects when low-paid individuals would receive money compared to high-paid individuals (Fig. 6a). Additionally, the vmPFC exhibited a similar pattern, showing increased activation when actions aimed to reduce inequity. Another research study found that this region reflects the outcome value for oneself when deciding personally and for the partner when making decisions on their behalf (Nicolle et al., 2012). Such findings suggest a strong consideration for equity that may surpass subjects' explicit statements.

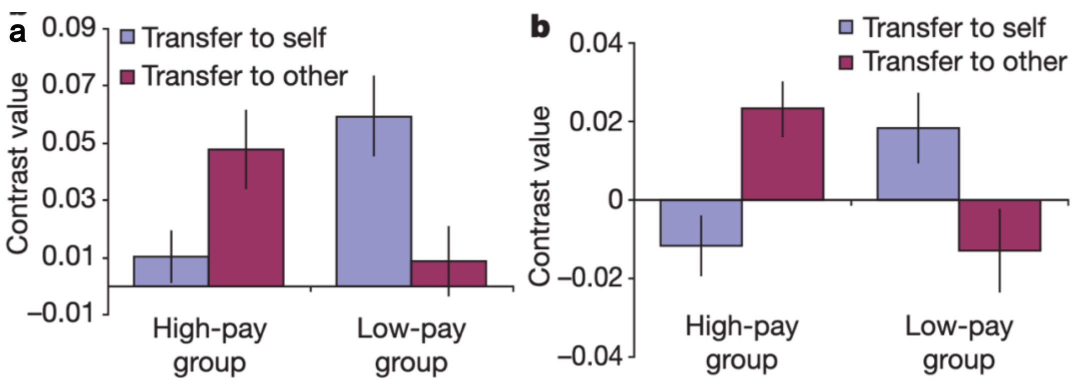

Fig. 6 Neural result (Tricomi et al., 2010)

In summary, this study demonstrates that both advantageous and disadvantageous inequities may be processed in the same brain region, suggesting the possibility of other factors moderating this strong fundamental equity consideration. This possibility was supported by Dawes (2012) who conducted the aforementioned psychology research, but this time combined with functional imaging. Using a similar paradigm, they found that both the vmPFC and the anterior insular, associated with empathy towards others' emotional or physical pain (Masten et al., 2011), were linked to decision-making, but only the latter was associated with revealed egalitarian preferences observed outside the scanner. Together, these functional imaging studies enhance our understanding of inequity aversion.

Another crucial aspect where functional imaging offers unique contributions, unattainable through psychology or lesion studies, is the exploration of how various concepts are represented in the brain, thereby unveiling our spontaneous thought patterns. For instance, employing Multivariate Pattern Analysis (MVPA) allows us to probe how mental states are organized.

In addressing this question, psychological studies have developed several theories concerning the organization of knowledge regarding mental states, encompassing dimensions like valence and arousal (Russell, 1980), warmth and competence (Fiske et al., 2002), agency and experience (Gray et al., 2007) among others. These theories reflect how we perceive others' minds across diverse situations. However, testing whether there exists a universal framework capable of organizing inner thoughts about different social constructs using the same dimensions poses a challenge.

For instance, Tamir et al. (2016) used 16 dimensions derived from psychological literature and an online survey to gauge the position of 166 mental states on each dimension. Subsequently, Principal Component Analysis (PCA) was employed to reduce the dimensions (Fig. 7). The resulting four principal components were termed rationality, social impact, human mind, and valence.

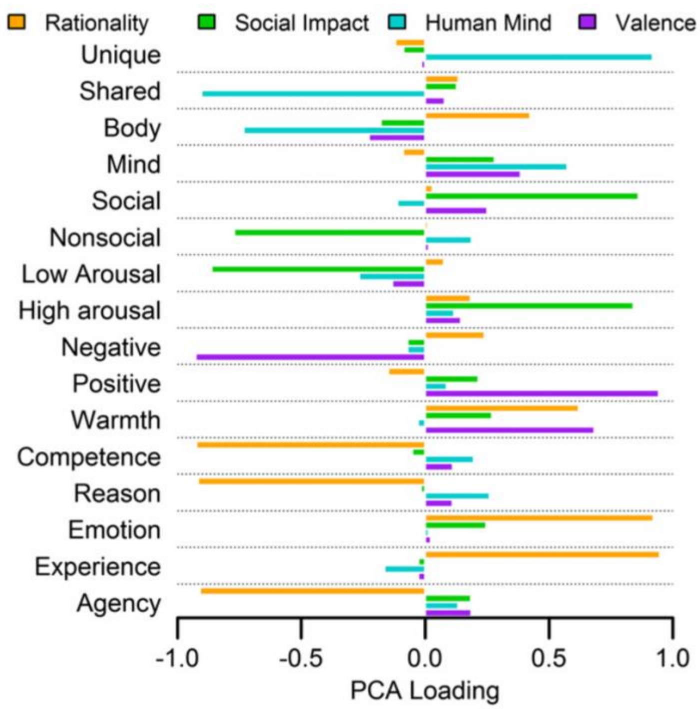

Fig. 7 PCA Loading of different mental states in several dimensions (Tamir et al., 2016)

During scanning, subjects were presented with 60 mental states across various trials and tasked with selecting scenarios that would evoke a particular mental state in another person. Neural activities corresponding to these mental states were analysed for pairwise similarity, which was then regressed onto the predictions of psychological similarity derived from PCA. The aim was to establish if mental states rated similarly on the PCA dimensions also exhibited similar neural activity patterns. The study found significant associations between three of the four dimensions (except human mind) and neural activities. These dimensions accounted for about one-third of the variance (R^2) in neural activities within the mental state network. Afterward, researchers divided R^2 by neural similarity and concluded that these three dimensions accounted for almost half of the variation in the neural representation of mental states.

Moreover, the authors assessed how conventional theories predict variations in neural activity. They discovered that five out of seven theories significantly accounted for variations in neural activities. However, the three dimensions explained nearly double the variance compared to the most successful original theories. This advanced functional imaging approach not only sheds light on longstanding debates within social psychology regarding the theory of mind but also facilitates the objective measurement of mental state activities.

Despite facing challenges like reverse inference (Poldrack, 2006), selective activation of brain regions solidifies our understanding. Additionally, functional imaging provides substantial objective continuous data in healthy subjects. Consequently, it enables the inference of different processes via distinct brain activity patterns or identical processes through similar brain activity patterns, aligning with Henson's (2005) argument of "Function-to-structure deduction" and "Structure-to-function induction." Ultimately, studying "where" in the brain allows us to infer and test "how" processes unfold.

Reference

Andreoni, J., & Bernheim, B. D. (2009). Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects. *Econometrica*, *77*(5), 1607–1636. https://doi.org/10.3982/ECTA7384

Bolton, G. E., & Ockenfels, A. (2000). ERC: A theory of equity, reciprocity, and competition. *American Economic Review*, *90*(1), 166–193. https://doi.org/10.1257/aer.90.1.166

Boudon, R. (2003). Beyond rational choice theory. *Annual Review of Sociology*, *29*(1), 1–21. https://doi.org/10.1146/annurev.soc.29.010202.100213

Camerer, C., & Hua Ho, T. (1999). Experience-weighted attraction learning in normal form games. *Econometrica*, *67*(4), 827–874. https://doi.org/10.1111/1468-0262.00054

Chang, L. J., Doll, B. B., van 't Wout, M., Frank, M. J., & Sanfey, A. G. (2010). Seeing is believing: Trustworthiness as a dynamic belief. *Cognitive Psychology*, *61*(2), 87–105. https://doi.org/10.1016/j.cogpsych.2010.03.001

Dawes, C. T., Fowler, J. H., Johnson, T., McElreath, R., & Smirnov, O. (2007). Egalitarian motives in humans. *Nature*, *446*(7137), Article 7137. https://doi.org/10.1038/nature05651

Dawes, C. T., Loewen, P. J., Schreiber, D., Simmons, A. N., Flagan, T., McElreath, R., Bokemper, S. E., Fowler, J. H., & Paulus, M. P. (2012). Neural basis of egalitarian behavior. *Proceedings of the National Academy of Sciences*, *109*(17), 6479–6483. https://doi.org/10.1073/pnas.1118653109

Delgado, M. R., Frank, R. H., & Phelps, E. A. (2005). Perceptions of moral character modulate the neural systems of reward during the trust game. *Nature Neuroscience*, *8*(11), Article 11. https://doi.org/10.1038/nn1575

Evdokimov, P., & Garfagnini, U. (2022). Higher-order learning. *Experimental Economics*, *25*(4), 1234–1266. https://doi.org/10.1007/s10683-021-09743-6

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, *114*(3), 817–868. https://doi.org/10.1162/003355399556151

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, *82*(6), 878–902. https://doi.org/10.1037/0022-3514.82.6.878

Fosgaard, T. R., Hansen, L. G., & Piovesan, M. (2013). Separating will from grace: An experiment on conformity and awareness in cheating. *Journal of Economic Behavior & Organization*, *93*, 279–284. https://doi.org/10.1016/j.jebo.2013.03.027

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science*, *315*(5812), 619–619. https://doi.org/10.1126/science.1134475

Greene, J. D., & Paxton, J. M. (2009). Patterns of neural activity associated with honest and dishonest moral decisions. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(30), 12506–12511. https://doi.org/10.1073/pnas.0900152106

Henson, R. (2005). What can functional neuroimaging tell the experimental psychologist? *The Quarterly Journal of Experimental Psychology Section A*, *58*(2), 193–233. https://doi.org/10.1080/02724980443000502

Lee, V. K., & Harris, L. T. (2013). How social cognition can inform social decision making. *Frontiers in Neuroscience*, *7*. https://doi.org/10.3389/fnins.2013.00259

Masten, C. L., Morelli, S. A., & Eisenberger, N. I. (2011). An fMRI investigation of empathy for ‘social pain’ and subsequent prosocial behavior. *NeuroImage*, *55*(1), 381–388. https://doi.org/10.1016/j.neuroimage.2010.11.060

Nicolle, A., Klein-Flügge, M. C., Hunt, L. T., Vlaev, I., Dolan, R. J., & Behrens, T. E. J. (2012). An agent independent axis for executed and modeled choice in medial prefrontal cortex. *Neuron*, *75*(6), 1114–1121. https://doi.org/10.1016/j.neuron.2012.07.023

Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences*, *10*(2), 59–63. https://doi.org/10.1016/j.tics.2005.12.004

Rabin, M. (1993). Incorporating fairness into game theory and economics. *The American Economic Review*, *83*(5), 1281–1302.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161–1178. https://doi.org/10.1037/h0077714

Salmon, T. C. (2001). An evaluation of econometric models of adaptive learning. *Econometrica*, *69*(6), 1597–1628. https://doi.org/10.1111/1468-0262.00258

Schultz, W. (2002). Getting formal with dopamine and reward. *Neuron*, *36*(2), 241–263. https://doi.org/10.1016/S0896-6273(02)00967-4

Speer, S. P. H., Smidts, A., & Boksem, M. A. S. (2022). Cognitive control and dishonesty. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2022.06.005

Tamir, D. I., Thornton, M. A., Contreras, J. M., & Mitchell, J. P. (2016). Neural evidence that three dimensions organize mental state representation: Rationality, social impact, and valence. *Proceedings of the National Academy of Sciences*, *113*(1), 194–199. https://doi.org/10.1073/pnas.1511905112

Tricomi, E., Rangel, A., Camerer, C. F., & O'Doherty, J. P. (2010). Neural evidence for inequality-averse social preferences. *Nature*, *463*(7284), Article 7284. https://doi.org/10.1038/nature08785

van 't Wout, M., & Sanfey, A. G. (2008). Friend or foe: The effect of implicit trustworthiness judgments in social decision-making. *Cognition*, *108*(3), 796–803. https://doi.org/10.1016/j.cognition.2008.07.002

Wilcox, N. T. (2006). Theories of learning in games and heterogeneity bias. *Econometrica*, *74*(5), 1271–1292. https://doi.org/10.1111/j.1468-0262.2006.00704.x

Zhu, L., Jenkins, A. C., Set, E., Scabini, D., Knight, R. T., Chiu, P. H., King-Casas, B., & Hsu, M. (2014). Damage to dorsolateral prefrontal cortex affects tradeoffs between honesty and self-interest. *Nature Neuroscience*, *17*(10), Article 10. https://doi.org/10.1038/nn.3798

Zhu, L., Jiang, Y., Scabini, D., Knight, R. T., & Hsu, M. (2019). Patients with basal ganglia damage show preserved learning in an economic game. *Nature Communications*, *10*(1), Article 1. https://doi.org/10.1038/s41467-019-08766-1