Published in final edited form as:

Cortex. 2020 June ; 127: 221–230. doi:10.1016/j.cortex.2020.02.014.

Breaking human social decision making into multiple components and then putting them together again

Shinsuke Suzuki[*,1,2], **John P. O'Doherty**[3,4]

[1]Brain, Mind and Markets Laboratory, Department of Finance, Faculty of Business and Economics, The University of Melbourne, Parkville, Australia

[2]Frontier Research Institute for Interdisciplinary Sciences, Tohoku University, Sendai, Japan

[3]Division of the Humanities and Social Sciences, California Institute of Technology, Pasadena, USA

[4]Computation and Neural Systems, California Institute of Technology, Pasadena, USA

Abstract

Most of our waking time as human beings is spent interacting with other individuals. In order to make good decisions in this social milieu, it is often necessary to make inferences about the internal states, traits and intentions of others. Recently, some progress has been made to uncover the neural computations underlying human social decision-making by combining functional magnetic resonance neuroimaging (fMRI) with computational modeling of behavior. Modeling of behavioral data allows us to identify key computations necessary for decision-making and how these computations are integrated. Furthermore, by correlating these computational variables against neuroimaging data, it has become possible to elucidate where in the brain various computational variables are implemented during social decision making. Here we review the current state of knowledge in the domain of social computational neuroscience. Findings to date have emphasized that social decisions are driven by multiple computations that are conducted in parallel and which are implemented in distinct brain regions. We suggest that further progress is going to depend on identifying how and where such variables get integrated in order to yield a coherent behavioral output.

Keywords

social cognition; decision-making; model-based fMRI; reinforcement learning; computational neuroscience

1. Introduction

A fundamental question in neuroscience is how we make a decision. A popular framework developed in economics, psychology and machine-learning is called value-based decision-

[*]Corresponding should be addressed to Shinsuke Suzuki (shinsuke.szk@gmail.com).

Competing financial interests

The authors declare no competing financial interest.

making (Rangel et al., 2008). The framework posits that (i) our brain assigns a scalar quantity, *subjective value*, to each of the available options, then (ii) selects the option with the highest value, and finally (iii) updates values of the options based on experienced outcome (i.e., learning). Application of formal models of value-based decision-making (e.g., reinforcement learning) to behavioral and neuroimaging data has uncovered neural mechanisms underlying human decision-making (Daw & Doya, 2006; Glimcher & Rustichini, 2004; O'DOHERTY et al., 2007; Schultz et al., 1997): for instance, subjective value signals encoded in the medial prefrontal cortex (Chib et al., 2009; Kable & Glimcher, 2007; Lebreton et al., 2009; Suzuki et al., 2017) and learning signals (i.e., reward prediction error) encoded in the striatum (McClure et al., 2003; O'Doherty et al., 2004; Rutledge et al., 2010).

In the last decade, researchers have employed formal theoretical approaches from economics, game theory and machine-learning to understand the neural underpinnings of human social behavior with reference to the value-based decision-making framework (Behrens et al., 2009; Dunne & O'Doherty, 2013; Hackel & Amodio, 2018; Lee, 2008; Ruff & Fehr, 2014). Social decision-making is evidently very complex, as it often requires inference about hidden states such as another's intentions, state of mind, traits and/or predispositions. Indeed, accumulating evidence suggests that multiple forms of computation performed in distinct brain regions might underlie social decision-making (Behrens et al., 2009; Charpentier & O'Doherty, 2018; Dunne & O'Doherty, 2013; Joiner et al., 2017; Konovalov et al., 2018; Lee & Seo, 2016; Ruff & Fehr, 2014; Wittmann et al., 2018). In other words, to compute and update one's own overall values for available decision options in a social situation, one might need to integrate multiple computations about one's own individual preferences, one's preferences about the outcomes that others can receive, socially-specific inferences about others, domain-general inferences about the environment and so on. Yet, much less is known about how these multiple computations necessary for social decision-making are integrated in the human brain.

Here, in this review, we discuss these issues, while maintaining a focus on studies that extend the value-based decision-making framework to social behavior. We first outline simple extensions of this framework to the social domain: decision-making for others, in which consideration of other individuals' welfare works as a modulatory factor, and learning through observing others, in which others' choice and its consequence work as a source of learning; and then review a more complex expansion of this framework to the domain of strategic decision-making/learning.

2. Value-based decision-making for others

In our daily life, we make value-based decisions not only for our own interest but also for the benefit of other individuals (e.g., charitable donation involving decision-making about how resources are allocated between oneself and others). In such decision-making, an individual computes the value of each option by considering both self and others' reward outcomes in line with her social preference such as warm-glow, inequity-aversion and envy-aversion (Crockett et al., 2017; Fehr & Shmidt, 1999; Fehr & Camerer, 2007; Fukuda et al., 2019; Harbaugh et al., 2007; Hula et al., 2018; Sanfey et al., 2003; Takahashi et al., 2009),

suggesting that multiple types of information are represented in the brain to guide choice. One study (Hutcherson et al., 2015) examined simple decisions about different allocations of monetary reward between oneself and an anonymous partner. They found that choice behavior and reaction-times can be well-captured by a computational model called the Multi-attribute Drift-Diffusion Model (Ratcliff & McKoon, 2008). Furthermore, monetary reward for oneself and that for the anonymous partner were found to be encoded in the ventral striatum and temporoparietal junction (TPJ) respectively.

Another study (Hsu et al., 2008) investigated decision-making between different donation plans to two groups of children living in an orphanage in northern Uganda. Importantly, the plans differed in terms of efficiency (i.e., the overall amounts of money donated to the two groups) and inequity (i.e., the difference in the amounts donated to the two groups). They found that information about efficiency was represented in a region of dorsal striatum (the putamen) while information about inequity was encoded in insula.

Moreover, we sometimes make decisions on behalf of others (e.g., leadership decisions) (Edelson et al., 2018; Jung et al., 2013; Nicolle et al., 2012; Ogawa et al., 2018). One study (Edelson et al., 2018) examined decision-making to take the lead (i.e., make a choice on behalf of the group) or not (i.e., follow the majority's choice). They found that participants in their experiment tended to avoid assuming leadership, especially when the choice was difficult; and that patterns of connectivity among brain regions encoding task-relevant variables (e.g., choice difficulty, probability of leading, and so on) predicted individual differences in leadership decisions and self-reported leadership scores.

Another important component in social value-based decision-making is learning for others (i.e., learning about the consequence of one's own action for other individuals). Researchers have identified neural underpinnings of learning for others to attain monetary reward (Christopoulos & King-Casas, 2015; Lockwood et al., 2016), to avoid painful electric shock (Lockwood et al., 2019), and to reduce exposure to unpleasantly loud noise (Sul et al., 2015). Some of these studies (Lockwood et al., 2016; 2019; Sul et al., 2015) have consistently found that ventral striatum tracks learning signals, reward prediction errors, when learning for others and oneself (but see (Christopoulos & King-Casas, 2015) for counter evidence). On the other hand, prediction error signals specific to learning for others have been found in the thalamus/caudate (Lockwood et al., 2019) and vmPFC expanding to subgenual anterior cingulate cortex (sgACC) (Christopoulos & King-Casas, 2015; Lockwood et al., 2016).

It is worth noting that there is likely to be considerable variation in social preference across individuals (Fehr & Camerer, 2007; Lange et al., 1997). Some people appear to care a lot about others' payoffs, while others seem to care much more about their own payoff. Such individual differences modulate neural responses to fairness (Haruno & Frith, 2009), other-regarding values (Sul et al., 2015) and prediction errors about others' rewards (Christopoulos & King-Casas, 2015).

While we have so far emphasized contributions of anatomically distinct brain regions to self-regarding and other-regarding computations, recent studies suggest the existence of more

flexible representations in the ventral-dorsal axis of medial prefrontal cortex (Nicolle et al., 2012; Sul et al., 2015). Comparing decision-making for oneself and that on behalf of others, one study (Nicolle et al., 2012) demonstrated that vmPFC tracked "executed value" utilized for the current choice. That is, the brain region signaled self- and other-referential values, when making a choice for self and for others, respectively. They further found that dorsomedial prefrontal cortex (dmPFC) encoded "modeled value", which is not used for the current choice but could be internally simulated (i.e., self- and other-values, when making a choice for others and for self, respectively). Another study (Sul et al., 2015) revealed that social preference modulated a spatial gradient from vmPFC predominantly representing self-value to dmPFC encoding other-value. These studies together challenge the view that distinct and non-overlapping neural mechanisms are utilized for social and non-social inferences.

Most of the studies discussed above have investigated social interactions with anonymous others. However, several studies suggest that neural responses to social decision-making can be modulated as a function of the relationship between self and other such as social distance (Strombach et al., 2015), friendship (Fareri et al., 2015) and group membership (Hackel et al., 2017; Hein et al., 2016). For example, in the context of decision-making for others, willingness to pay for another individual's benefit declines with an increase in social distance, and social-distance-dependent choices were found to be associated with neural activity in TPJ and vmPFC and in the functional coupling between those areas (Strombach et al., 2015).

3. Value-based decision-making through observing others

To make appropriate decisions, one needs to learn the values of available actions (options) and other features of the environment. Such learning can be accomplished not only by one's own experience but also by observing another individual's experience. This type of observational learning exists in multiple species, and has been directly tested in a range of species from rodents to humans (Burke et al., 2010; Cooper et al., 2011; Hill et al., 2016; Zentall, 2012). This form of learning is likely to be beneficial for survival as it enables individuals to efficiently acquire knowledge about the world without direct experience.

Recent human neuroimaging studies suggest that, for observational learning, human observers utilize two sources of information in order to acquire knowledge from an observee: the rewards obtained by the observee in relation to particular choices, and the actions performed by the observee (Burke et al., 2010; Suzuki et al., 2012). More precisely, prediction errors about the reward outcomes received by the observee have been found in vmPFC and dorsal striatum, and these may be used to update the value of the option chosen by the observee, in a similar manner to that which occurs during conventional reinforcement learning through direct experience. Indeed, a meta-analysis found that the vmPFC tracks reward prediction errors about both experienced and observed outcomes (Morelli et al., 2015). On the other hand, prediction errors about the observee's actions (i.e., the discrepancy between the observee's actual choice and the observer's prediction of the choice) have been found to be encoded in the dorsolateral prefrontal cortex (dlPFC) as well as other brain structures such as dorsomedial prefrontal cortex (dmPFC) and inferior parietal

lobule (IPL) (Burke et al., 2010; Suzuki et al., 2012). Moreover, another line of studies focused on fear conditioning and highlighted the pivotal role of amygdala in both learning from experienced and observed outcomes (Olsson et al., 2007; Olsson & Phelps, 2007).

In decision-making and learning through direct experience, there exist two key strategies (Balleine & O'Doherty, 2010): one is a goal-directed strategy that tracks causal relations between actions and outcomes, and the other is a habitual strategy in which actions are automatically elicited by environmental states. Interestingly, a similar dichotomy between goal-directed and habitual strategies has been suggested to apply when learning through observation (Liljeholm et al., 2012). However, the underlying neural mechanisms of these two strategies still remain elusive (Dunne et al., 2016; Liljeholm et al., 2012).

Learning about another individual's reliability or trustworthiness and competence through observing her behavior is also useful, especially when making a decision about whether or not to take into account her advice (Behrens et al., 2008; Boorman et al., 2013). One study (Behrens et al., 2008) examined a case, in which for optimal decisions participants were required to combine learning about option values through direct reward feedback and learning about the adviser's reliability through observation. They found that these two types of learning were formed in parallel in the brain: with the ventral striatum tracking learning about value from reward feedback, and the right posterior superior temporal sulcus (pSTS) and TPJ tracking learning about the adviser's reliability. Furthermore, neural signatures of uncertainty in the two types of learning were found in distinct sub-regions of dACC, consistent with the postulation that uncertainty of the estimation modulates the speed of learning (Behrens et al., 2007). Another study (Wittmann et al., 2016) used a mini-game that requires participants to estimate their own and other players' ability in cooperative and competitive contexts. The results of that study showed that the abilities of the participants themselves on the task and those of the other players were estimated based on past performance, and were found to be represented in the vmPFC and dmPFC respectively. In addition to reliability and ability, researchers have assessed social learning about other types of traits (Delgado et al., 2005; Hackel et al., 2015; Stanley, 2016). For example, Hackel et al. (Hackel et al., 2015) devised a task that allows one to dissociate learning about others' generosity from learning about the reward obtained from others. Utilizing this task, they revealed that, while both types of learning recruited ventral striatum, learning about others' generosity specifically employed a network of brain regions associated with social cognition (e.g., TPJ and precuneus).

More broadly, neural signatures of learning signals (i.e., prediction errors) have been reported in many other social situations. For example, in the case of learning about ownership, prediction errors about others' and self ownership are represented in distinct sub-regions of dACC along the antero-posterior axis: the anterior part tracks prediction errors about whether objects belong to others, while the posterior part tracks prediction errors about individuals' own ownership (Lockwood et al., 2018). Furthermore, another study examined a teacher-student interaction in which the teacher informed the student whether the student's choice was rewarding or not (Apps et al., 2016). Those authors demonstrated a role for anterior dACC in the teacher's brain in signaling prediction errors about the student's rewards, suggesting that teachers vicariously kept track of their students' learning

progress. Importantly, these prediction error signals in ACC cannot be attributed to the domain-general process of error/conflict detection (Botvinick et al., 2004), because the signals were observed only in a particular condition (Lockwood et al., 2018), and were significant after controlling for effects of error trials and surprise (i.e., unsigned prediction error) (Apps et al., 2015).

Apart from value-based decision-making, information provided by others would be useful for forming and updating one's belief about oneself. One study (Will et al., 2017) examined how appraisals from others shape our self-esteem. They found that fluctuation in self-esteem was driven by a prediction error corresponding to the discrepancy between expected and received social feedback. This was in turn represented in the ventral striatum and sgACC.

4. Value-based decision-making in strategic interactions

In the real world, we are often engaged in bilateral or reciprocal interactions, in which an individual needs to take into account predictions about another agent's intentions in order to make an advantageous decision, while the other agent also strives to predict the individual's intentions at the same time (Camerer, 2003). Researchers have begun to uncover computational processes for such strategic decision-making/learning and its neural underpinnings (Fareri, Chang, & Delgado, 2015; Hampton, Bossaerts, & O'Doherty, 2008; Haruno & Kawato, 2009; Hill et al., 2017; Lee & Seo, 2016; Suzuki, Adachi, Dunne, Bossaerts, & O'Doherty, 2015; Xiang, Ray, Lohrenz, Dayan, & Montague, 2012; Yoshida, Seymour, Friston, & Dolan, 2010; Zhu, Mathewson, & Hsu, 2012).

One important form of strategic decision-making arises when it is necessary to form a coordination or consensus within a group. One study (Suzuki et al., 2015) devised a novel experimental task in which in the main condition participants tried to make a unanimous consensus with other human participants. Behavioral modeling together with analysis of neuroimaging data demonstrated that one's decisions were guided by three separate factors: knowledge about one's own preference, information about the prior choices made by the majority of group-members, as well as an inference about how much each option is doggedly stuck to by the other group-members. These three different variables were found to be represented in the vmPFC, TPJ and IPL respectively (Fig. 1AB). Note that TPJ and IPL activations sometimes overlap, but this was not the case in this study (the peak MNI coordinates: [60 −46 10] for TPJ and [30 −52 34] for IPL). Importantly, the experimental task used in this study had a control condition in which participants interacted with computer agents programmed to mimic actual human behavior. Comparison of the main and the control conditions revealed a significant difference in the neural representation of group-members' prior choices in TPJ, but not for the other variables. This suggests that information about the group-members' prior choices is processed in a social-specific manner in TPJ, while information about one's own preference and inference about the stickiness of an option (i.e., how much each option is stuck to by the other human group-members or computer agents) are processed in a domain-general manner in vmPFC and IPL respectively.

Another study (Hampton et al., 2008) examined a competitive interaction between two individuals, by using an experimental task originally developed in economics. The task,

called the *inspection game*, models repeated interactions between an employee and her employer, in which the employee decides to work or to shirk (i.e., not to work) while the employer decides to inspect or not to inspect her employee. Each player gets a higher payoff if they can outsmart the opponent. For example, an employee obtains a higher reward by shirking without being inspected by the employer. The authors identified two computations related to forming a prediction about the opponent's next move. One is learning from the opponent's past choices, driven by prediction error, which was found to be encoded in the ventral striatum. The other is higher-order reasoning about how one's own current move will influence the opponent's next choice, which was found to be encoded in pSTS/TPJ. It is also worth noting that a recent study combining computational modeling and neuroimaging with non-invasive brain stimulation largely replicated the original findings of the Hampton et al. study, while establishing that the TPJ signal is causally relevant for computing a higher-order inference about the influence of one's own action on the opponent's next choice (Hill et al., 2017). In that study, theta burst stimulation over the TPJ which temporarily disrupts the functions of that region was found to result in a reduced tendency to engage the higher order inference compared to a sham control condition.

Complex strategic interactions often involve deep recursive reasoning about another's mental state: inference about your inference about my inference about your inference and so on (Camerer, 2003). A cognitive hierarchy theory, originally developed in game theory, posits that an individual conducts recursive inferences to a one-step higher level of depth than one's opponent in order to gain an advantage over the opponent (Camerer, 2003). In other words, a reasonable strategy is to estimate the opponent's depth-of-reasoning and adjust one's own behavior so as to be tailored to that estimation (by being one step more sophisticated). Although in general this type of inference is itself computationally costly, several formal models based on Bayesian inference have been proposed (Ray et al., 2008; Yoshida et al., 2008). Correlating these models with fMRI signals, Yoshida et al. showed that uncertainty about the estimation of the opponent's depth-of-reasoning was represented in dmPFC, while dlPFC tracked the depth of one's own strategy (Yoshida et al., 2010). Furthermore, Xiang et al. demonstrated differential brain regions encoded reward prediction error signals with respect to individual difference in participants' depth-of-reasoning (Xiang et al., 2012).

Note that complex strategic decision-making is not mutually exclusive from the issues discussed in the previous sections. For example, inference about the opponent's depth-of-reasoning (Yoshida et al., 2010) can be interpreted as a form of learning about her traits (see the section of value-based decision-making through observing others). Furthermore, decision-making in a strategic game called the Ultimatum game has been found to be affected by one's preference for fairness (see the section of value-based decision-making for others) (Chang & Sanfey, 2013; Falk et al., 2003; Sanfey et al., 2003).

5. Integration of multiple computations in social value-based decision-making

We have considered evidence supporting the possibility that multiple computational strategies are involved in parallel during many different forms of social value-based decision-making. However, the different computations need to be integrated somehow in order to generate a coherent behavioral output. In our review so far we have revealed that neuroimaging studies have identified a number of different forms of computation during social-decision-making which are represented in discrete brain regions. In most of the computational models mentioned above, the form of the integration of the variables is to compute a subjective value or choice probability for a given option. For example, in models of decision-making for others (e.g., Fukuda et al., 2019; Hutcherson et al., 2015), a subjective value for each option is computed by integrating information about the amount of reward delivered to oneself and others (Fig 2A). In the contexts of observational learning (e.g., Burke et al., 2010; Suzuki et al., 2012), leaning signals from others' choices and reward outcomes are integrated to compute the value of each option (Fig 2B). In strategic decision-making models (e.g., Hampton et al., 2008; Hill et al., 2017; Suzuki et al., 2015), value computation requires the integration of multiple types of inference (Fig 1 and 2C). Given these model structures (Figs 1 and 2), we suggest that a brain region engaged in the integration process must (1) encode the integrated subjective value or choice probability signals assigned by the computational model (Fig. 1C) and (2) have functional connectivity with regions encoding each of the individual key computational variables (Fig. 1D). In other words, if a brain region is implicated in the information integration, the region must satisfy the above two criteria.

The first criterion has been examined in many studies. Correlating fMRI signals with the model-derived overall subjective value or choice probability signals, convergent evidence suggests that key computational variables necessary for decision-making are integrated in the vmPFC including the rostral ACC (rACC) and/or the dmPFC including dACC (e.g., Behrens et al., 2008; Hsu et al., 2008; Hutcherson et al., 2015; Suzuki et al., 2012).

On the other hand, to our knowledge, only a few studies (Fukuda et al., 2019; Hampton et al., 2008; Hill et al., 2017; Suzuki et al., 2015) have examined both of the two criteria. For example, as mentioned in the previous section, Suzuki et al. (Suzuki et al., 2015) first identified three key computational variables and their neural correlates for consensus formation in a group: vmPFC encoding one's own preference for each of the available options, TPJ encoding group-members' prior choices and IPL encoding one's inference about how much each option was stuck to by the other group-members (Fig. 1AB). Next, they found two brain regions, rACC and posterior dACC, that satisfied the first criterion: that is, fMRI signals in these two regions were correlated with modeled choice probability derived by integration of the three key computational variables (Fig. 1C). Finally, to examine the second criterion, a functional connectivity analysis was employed (Psycho-Physiological Interaction analysis (Friston et al., 1997)) which demonstrated that the posterior dACC, but not the rACC, had increased connectivity at the time of decision with each of the three regions, the vmPFC, the TPJ, and the IPL, that individually tracked the three key

computational variables (Fig. 1D). Taken together, only the posterior dACC satisfies both of the two criteria, suggesting that the three key computational variables involved in consensus decision-making are integrated in posterior dACC.

In the context of competitive interactions, Hampton et al. suggest that integration process occurs in dmPFC (Hampton et al., 2008). The dmPFC was found to represent overall subjective value and have functional connectivity with the other regions, the ventral striatum and the pSTS/TPJ, responsible for individual computations underlying the decision-making (i.e., learning from the opponent's past choices and higher-order reasoning about how one's own choice will influence the opponent's next choice). This account of integration is further supported by a recent study showing that disruption of the TPJ alters its functional connectivity with the dmPFC (Hill et al., 2017).

In the context of decision-making for others, one study (Fukuda et al., 2019) tested for areas meeting the two criteria, by combining two types of connectivity analysis: Psycho-Physiological Interaction analysis, PPI (Friston et al., 1997), and Dynamical Causal Modeling, DCM (Friston et al., 2003). Note that PPI is based on a regression model and thus cannot examine directionality of the connectivity, while DCM is based on a model of causal interactions of brain regions and can thus enable inference about the directionality of the effects. In Fukuda et al., these two connectivity analysis approaches consistently showed that integration of information about self- and others' rewards occurred in the vmPFC, including the adjacent rACC (Fukuda et al., 2019).

Furthermore, some other studies have aimed to address the issue of integration by using connectivity analyses, although they did not directly test the above two criteria. For example, van den Bos et al. suggest that multiple computations necessary for bidding behavior in an auction are integrated in vmPFC and striatum (van den Bos et al., 2013). Smith et al. proposed that the value of social stimuli (i.e., attractiveness of others' faces) is computed in the vmPFC via interactions with other regions such as the TPJ and middle temporal gyrus (Smith et al., 2014). Finally, based on a meta-analytic connectivity analysis evaluating co-activation patterns across various tasks, the authors of one study (Alcalá-López et al., 2017) concluded that diverse neural circuits for from low-level sensory to high-level associative processes mediate human social cognitive capacities.

These findings could, we believe, provide a possible account for HOW social information is integrated with simple non-social decision-making processes. In studies on simple decision-making, it has been suggested that values of available options and goals in the vmPFC (including rACC) are utilized as inputs for computing values for actions in the posterior dmPFC (posterior dACC), and then finally transformed into a motor command in the motor cortex (Hare et al., 2011) . The findings obtained in Suzuki et al., Hampton et al. and Fukuda et al. could suggest that, in the contexts of strategic decision-making, social information modulates the basic decision-making process at the stage of action value computation in the posterior dmPFC, while social information operates on the upstream stage (i.e., the computation of option and goal values in the vmPFC) in the context of decisions for others. This account further motivates another fascinating question: how is anterior dmPFC (anterior dACC), located between vmPFC and posterior dmPFC along the rostral-caudal axis

of medial prefrontal cortex, involved in the social decision-making process? This question is of particular importance as the anterior dmPFC has been proposed to play a central role in social cognition (Amodio & Frith, 2006; Apps et al., 2016). Given the sparsity of studies to date that have addressed both of the above two criteria for how integration might happen across different computational strategies, however, future studies will need to address how and where discrete social computations are integrated across a wide array of different task domains and computational variables.

6. Conclusions

It is widely believed that decisions in social contexts are made through integrating multiple types of inference about one's own rewards, others' rewards, others' mental-states and so on. In the last decade, the notion has been supported by a computational modeling approach combined with neuroimaging (Behrens et al., 2009; Charpentier & O'Doherty, 2018; Dunne & O'Doherty, 2013; Joiner et al., 2017; Konovalov et al., 2018; Lee & Seo, 2016; Wittmann et al., 2018). Construction of a formal model that can account for behavioral data enables us to identify key variables necessary for decision-making, providing significant insights into what computations underlie human social behavior. Furthermore, correlating the key variables with neuroimaging data informs us where in the brain these computations are implemented.

On the other hand, a more challenging and less explored issue is how these computations are integrated in the brain to guide our social behavior. In a broader sense, this issue is related to a long-lasting question in neuroscience, known as "The Binding Problem" (Roskies, 1999). While in this review we have introduced some studies addressing this issue, more evidence is needed for a more comprehensive understanding of the information integration process. For example, to understand information integration process in the brain, it would also be essential to examine the nature of causal interactions (i.e., the direction of information flow) among multiple brain regions, which cannot be tested by correlation-based connectivity analysis methods such as psychophysiological interaction analysis (Friston et al., 1997). An interesting and important avenue in social decision-making would be to utilize Dynamic Causal Modeling or/and non-invasive brain stimulation together with computational modeling (Hein, Morishima, et al., 2016; Hill et al., 2017).

Important issues we have not addressed in this review are the perceptual aspects of social decision-making. Perception of social stimuli, especially others' faces, plays an important role in real-world decisions such as mate choice (Fletcher et al., 1999), electoral behavior (Todorov et al., 2005) and sentencing judgments (Blair et al., 2004). While several brain regions have been found to represent others' attractiveness (O'Doherty et al., 2003), emotion (Wegrzyn et al., 2015) and trustworthiness (Todorov et al., 2008; Winston et al., 2002) as perceived from their faces, much less is known about how these perceptions are constructed from low-level visual inputs. Another line of studies has examined neural mechanisms underlying perception of animacy or biological motion (Giese & Poggio, 2003; Schultz & Bülthoff, 2019). Such studies have implicated a brain network including pSTS in detection of animacy of abstract stimuli (e.g., moving dots), but the underlying computations still remain elusive. Future studies could fruitfully provide neurocomputational accounts that

bridge low-level sensory inputs and the higher-order perception of social stimuli (Chang & Tsao, 2017; Lin et al., 2019; Oosterhof & Todorov, 2008).

To conclude, in this review, we discuss recent advancement in the studies of human social value-based decision-making. Despite the consensus that multiple types of computations underlie the decision-making, our understanding of how these computations are integrated to guide behavior is still in its infancy. Further elucidation of the integration process by combining neuroimaging, brain stimulation, computational modeling and connectivity analyses would be a critical step towards a more comprehensive understanding of human social decision-making.

Acknowledgments

This work was supported by the JSPS KAKENHI Grants JP17H05933 and JP17H06022 (S.S.) and the NIMH Caltech Conte Center for the Neurobiology of Social Decision Making (J.P.O).

References

Alcalá-López D, Smallwood J, Jefferies E, Overwalle F, Vogeley K, Mars RB, Turetsky BI, Laird AR, Fox PT, Eickhoff SB, & Bzdok D (2017). Computing the Social Brain Connectome Across Systems and States. Cerebral Cortex, 28(7), 2207–2232. 10.1093/cercor/bhx121

Amodio DM, & Frith CD (2006). Meeting of minds: the medial frontal cortex and social cognition. Nature Reviews Neuroscience, 7(4), nrn1884 10.1038/nrn1884

Apps M, Lesage E, & Ramnani N (2015). Vicarious Reinforcement Learning Signals When Instructing Others. The Journal of Neuroscience, 35(7), 2904–2913. 10.1523/jneurosci.3669-14.2015 [PubMed: 25698730]

Apps M, Rushworth M, & Chang S (2016). The Anterior Cingulate Gyrus and Social Cognition: Tracking the Motivation of Others. Neuron, 90(4), 692–707. 10.1016/j.neuron.2016.04.018 [PubMed: 27196973]

Balleine BW, & O'Doherty JP (2010). Human and Rodent Homologies in Action Control: Corticostriatal Determinants of Goal-Directed and Habitual Action. Neuropsychopharmacology, 35(1), 48–69. 10.1038/npp.2009.131 [PubMed: 19776734]

Behrens TE, Hunt LT, & Rushworth MF (2009). The Computation of Social Behavior. Science, 324(5931), 1160–1164. 10.1126/science.1169694 [PubMed: 19478175]

Behrens TE, Hunt LT, Woolrich MW, & Rushworth MF (2008). Associative learning of social value. Nature, 456(7219), 245 10.1038/nature07538 [PubMed: 19005555]

Behrens TE, Woolrich MW, Walton ME, & Rushworth MF (2007). Learning the value of information in an uncertain world. Nature Neuroscience, 10(9), 1214–1221. 10.1038/nn1954 [PubMed: 17676057]

Blair I, Judd C, & Chapleau K (2004). The Influence of Afrocentric Facial Features in Criminal Sentencing. Psychological Science, 15(10), 674–679. 10.1111/j.0956-7976.2004.00739.x [PubMed: 15447638]

Boorman ED, O'Doherty JP, Adolphs R, & Rangel A (2013). The Behavioral and Neural Mechanisms Underlying the Tracking of Expertise. Neuron, 80(6), 1558–1571. 10.1016/j.neuron.2013.10.024 [PubMed: 24360551]

Botvinick MM, Cohen JD, & Carter CS (2004). Conflict monitoring and anterior cingulate cortex: an update. Trends in Cognitive Sciences, 8(12), 539–546. 10.1016/j.tics.2004.10.003 [PubMed: 15556023]

Burke CJ, Tobler PN, Baddeley M, & Schultz W (2010). Neural mechanisms of observational learning. Proceedings of the National Academy of Sciences, 107(32), 14431–14436. 10.1073/pnas.1003111107

Camerer CF (2003). Behavioral Game Theory: Experiments in Strategic Interaction.

Camerer C, Ho HT, & Chong KJ (2004). A Cognitive Hierarchy Model of Games. The Quarterly Journal of Economics, 119(3), 861–898. 10.1162/0033553041502225

Chang LJ, & Sanfey AG (2013). Great expectations: neural computations underlying the use of social norms in decision-making. Social Cognitive and Affective Neuroscience, 8(3), 277–284. 10.1093/scan/nsr094 [PubMed: 22198968]

Chang L, & Tsao DY (2017). The Code for Facial Identity in the Primate Brain. Cell, 169(6), 1013–1028.e14. 10.1016/j.cell.2017.05.011 [PubMed: 28575666]

Charpentier CJ, & O'Doherty JP (2018). The application of computational models to social neuroscience: promises and pitfalls. Social Neuroscience, 13(6), 637–647. 10.1080/17470919.2018.1518834 [PubMed: 30173633]

Chib VS, Rangel A, Shimojo S, & O'Doherty JP (2009). Evidence for a Common Representation of Decision Values for Dissimilar Goods in Human Ventromedial Prefrontal Cortex. The Journal of Neuroscience, 29(39), 12315–12320. 10.1523/jneurosci.2575-09.2009 [PubMed: 19793990]

Christopoulos GI, & King-Casas B (2015). With you or against you: Social orientation dependent learning signals guide actions made for others. NeuroImage, 104, 326–335. 10.1016/j.neuroimage.2014.09.011 [PubMed: 25224998]

Cooper JC, Dunne S, Furey T, & O'Doherty JP (2011). Human Dorsal Striatum Encodes Prediction Errors during Observational Learning of Instrumental Actions. Journal of Cognitive Neuroscience, 24(1), 106–118. 10.1162/jocn_a_00114 [PubMed: 21812568]

Crockett MJ, Siegel JZ, Kurth-Nelson Z, Dayan P, & Dolan RJ (2017). Moral transgressions corrupt neural representations of value. Nature Neuroscience, 20(6), 879 10.1038/nn.4557 [PubMed: 28459442]

Daw ND, & Doya K (2006). The computational neurobiology of learning and reward. Current Opinion in Neurobiology, 16(2), 199–204. 10.1016/j.conb.2006.03.006 [PubMed: 16563737]

Delgado M, Frank R, & Phelps E (2005). Perceptions of moral character modulate the neural systems of reward during the trust game. Nature Neuroscience, 8(11), 1611–1618. 10.1038/nn1575 [PubMed: 16222226]

Dunne S, D'Souza A, & O'Doherty JP (2016). The involvement of model-based but not model-free learning signals during observational reward learning in the absence of choice. Journal of Neurophysiology, 115(6), 3195–3203. 10.1152/jn.00046.2016 [PubMed: 27052578]

Dunne S, & O'Doherty JP (2013). Insights from the application of computational neuroimaging to social neuroscience. Current Opinion in Neurobiology, 23(3), 387–392. 10.1016/j.conb.2013.02.007 [PubMed: 23518140]

Edelson MG, Polania R, Ruff CC, Fehr E, & Hare TA (2018). Computational and neurobiological foundations of leadership decisions. Science, 361(6401), eaat0036 10.1126/science.aat0036 [PubMed: 30072510]

Falk A, Fehr E, & Fischbacher U (2003). On the Nature of Fair Behavior. Economic Inquiry, 41(1), 20–26. 10.1093/ei/41.1.20

Fareri DS, Chang LJ, & Delgado MR (2015). Computational Substrates of Social Value in Interpersonal Collaboration. The Journal of Neuroscience, 35(21), 8170–8180. 10.1523/jneurosci.4775-14.2015 [PubMed: 26019333]

Fehr E, & Shmidt K (1999). A Theory of Fairness, Competition, and Cooperation. The Quarterly Journal of Economics, 114(3), 817–868. 10.1162/003355399556151

Fehr E, & Camerer CF (2007). Social neuroeconomics: the neural circuitry of social preferences. Trends in Cognitive Sciences, 11(10), 419–427. 10.1016/j.tics.2007.09.002 [PubMed: 17913566]

Fletcher GJ, Simpson JA, Thomas G, & Giles L (1999). Ideals in intimate relationships. Journal of Personality and Social Psychology, 76(1), 72–89. 10.1037/0022-3514.76.1.72 [PubMed: 9972554]

Friston K ., Buechel C, Fink G ., Morris J, Rolls E, & Dolan R . (1997). Psychophysiological and Modulatory Interactions in Neuroimaging. NeuroImage, 6(3), 218–229. 10.1006/nimg.1997.0291 [PubMed: 9344826]

Friston KJ, Harrison L, & Penny W (2003). Dynamic causal modelling. NeuroImage, 19(4), 1273–1302. 10.1016/s1053-8119(03)00202-7 [PubMed: 12948688]

Fukuda H, Ma N, Suzuki S, Harasawa N, Ueno K, Gardner JL, Ichinohe N, Haruno M, Cheng K, & Nakahara H (2019). Computing Social Value Conversion in the Human Brain. Journal of Neuroscience, 3117–3118. 10.1523/jneurosci.3117-18.2019

Giese MA, & Poggio T (2003). Neural mechanisms for the recognition of biological movements. Nature Reviews Neuroscience, 4(3), 179–192. 10.1038/nrn1057 [PubMed: 12612631]

Glimcher PW, & Rustichini A (2004). Neuroeconomics: The Consilience of Brain and Decision. Science, 306(5695), 447–452. 10.1126/science.1102566 [PubMed: 15486291]

Hackel LM, & Amodio DM (2018). Computational neuroscience approaches to social cognition. Current Opinion in Psychology, 24(Psychol. Sci. 15 2004), 92–97. 10.1016/j.copsyc.2018.09.001 [PubMed: 30388495]

Hackel LM, Doll BB, & Amodio DM (2015). Instrumental learning of traits versus rewards: dissociable neural correlates and effects on choice. Nature Neuroscience, 18(9), 1233–1235. 10.1038/nn.4080 [PubMed: 26237363]

Hackel LM, Zaki J, & Bavel JJ (2017). Social Identity Shapes Social Valuation: Evidence from Prosocial Behavior and Vicarious Reward. Social Cognitive and Affective Neuroscience, 12(8), nsx045-. 10.1093/scan/nsx045

Hampton AN, Bossaerts P, & O'Doherty JP (2008). Neural correlates of mentalizing-related computations during strategic interactions in humans. Proceedings of the National Academy of Sciences, 105(18), 6741–6746. 10.1073/pnas.0711099105

Harbaugh WT, Mayr U, & Burghart DR (2007). Neural Responses to Taxation and Voluntary Giving Reveal Motives for Charitable Donations. Science, 316(5831), 1622–1625. 10.1126/science.1140738 [PubMed: 17569866]

Hare TA, Schultz W, Camerer CF, O'Doherty JP, & Rangel A (2011). Transformation of stimulus value signals into motor commands during simple choice. Proceedings of the National Academy of Sciences, 108(44), 18120–18125. 10.1073/pnas.1109322108

Haruno M, & Frith CD (2009). Activity in the amygdala elicited by unfair divisions predicts social value orientation. Nature Neuroscience, 13(2), 160–161. 10.1038/nn.2468 [PubMed: 20023652]

Haruno M, & Kawato M (2009). Activity in the Superior Temporal Sulcus Highlights Learning Competence in an Interaction Game. The Journal of Neuroscience, 29(14), 4542–4547. 10.1523/jneurosci.2707-08.2009 [PubMed: 19357279]

Hein G, Engelmann JB, Vollberg MC, & Tobler PN (2016). How learning shapes the empathic brain. Proceedings of the National Academy of Sciences, 113(1), 80–85. 10.1073/pnas.1514539112

Hein G, Morishima Y, Leiberg S, Sul S, & Fehr E (2016). The brain's functional network architecture reveals human motives. Science, 351(6277), 1074–1078. 10.1126/science.aac7992 [PubMed: 26941317]

Hill CA, Suzuki S, Polania R, Moisa M, O'Doherty JP, & Ruff CC (2017). A causal account of the brain network computations underlying strategic social behavior. Nature Neuroscience, 20(8), 1142–1149. 10.1038/nn.4602 [PubMed: 28692061]

Hill MR, Boorman ED, & Fried I (2016). Observational learning computations in neurons of the human anterior cingulate cortex. Nature Communications, 7(1), 12722 10.1038/ncomms12722

Hsu M, Anen C, & Quartz SR (2008). The Right and the Good: Distributive Justice and Neural Encoding of Equity and Efficiency. Science, 320(5879), 1092–1095. 10.1126/science.1153651 [PubMed: 18467558]

Hula A, Vilares I, Lohrenz T, Dayan P, & Montague PR (2018). A model of risk and mental state shifts during social interaction. PLOS Computational Biology, 14(2), e1005935 10.1371/journal.pcbi.1005935 [PubMed: 29447153]

Hutcherson CA, Bushong B, & Rangel A (2015). A Neurocomputational Model of Altruistic Choice and Its Implications. Neuron, 87(2), 451–462. 10.1016/j.neuron.2015.06.031 [PubMed: 26182424]

Joiner J, Piva M, Turrin C, & Chang SW (2017). Social learning through prediction error in the brain. Npj Science of Learning, 2(1), 8 10.1038/s41539-017-0009-2 [PubMed: 30631454]

Jung D, Sul S, & Kim H (2013). Dissociable neural processes underlying risky decisions for self versus other. Frontiers in Neuroscience, 7, 15 10.3389/fnins.2013.00015 [PubMed: 23519016]

Kable JW, & Glimcher PW (2007). The neural correlates of subjective value during intertemporal choice. Nature Neuroscience, 10(12), nn2007 10.1038/nn2007

Konovalov A, Hu J, & Ruff CC (2018). Neurocomputational Approaches to Social Behavior. Current Opinion in Psychology. 10.1016/j.copsyc.2018.04.009

Lange PA, Bruin EM, Otten W, & Joireman JA (1997). Development of prosocial, individualistic, and competitive orientations: Theory and preliminary evidence. Journal of Personality and Social Psychology, 73(4), 733 10.1037/0022-3514.73.4.733 [PubMed: 9325591]

Lebreton M, Jorge S, Michel V, Thirion B, & Pessiglione M (2009). An Automatic Valuation System in the Human Brain: Evidence from Functional Neuroimaging. Neuron, 64(3), 431–439. 10.1016/j.neuron.2009.09.040 [PubMed: 19914190]

Lee D (2008). Game theory and neural basis of social decision making. Nature Neuroscience, 11(4), 404–409. 10.1038/nn2065 [PubMed: 18368047]

Lee D, & Seo H (2016). Neural Basis of Strategic Decision Making. Trends in Neurosciences, 39(1), 40–48. 10.1016/j.tins.2015.11.002 [PubMed: 26688301]

Liljeholm M, Molloy CJ, & O'Doherty JP (2012). Dissociable Brain Systems Mediate Vicarious Learning of Stimulus–Response and Action–Outcome Contingencies. The Journal of Neuroscience, 32(29), 9878–9886. 10.1523/jneurosci.0548-12.2012 [PubMed: 22815503]

Lin C, Keles U, & Adolphs R (2019). Comprehensive trait attributions show that face impressions are organized in four dimensions. 10.31234/osf.io/87nex

Lockwood PL, Apps MA, Valton V, Viding E, & Roiser JP (2016). Neurocomputational mechanisms of prosocial learning and links to empathy. Proceedings of the National Academy of Sciences, 113(35), 9763–9768. 10.1073/pnas.1603198113

Lockwood PL, Klein-Flügge M, Abdurahman A, & Crockett MJ (2019). Neural signatures of model-free learning when avoiding harm to self and other. BioRxiv, 718106 10.1101/718106

Lockwood PL, Wittmann MK, Apps MA, Klein-Flügge MC, Crockett MJ, Humphreys GW, & Rushworth MF (2018). Neural mechanisms for learning self and other ownership. Nature Communications, 9(1), 4747 10.1038/s41467-018-07231-9

McClure SM, Berns GS, & Montague PR (2003). Temporal Prediction Errors in a Passive Learning Task Activate Human Striatum. Neuron, 38(2), 339–346. 10.1016/s0896-6273(03)00154-5 [PubMed: 12718866]

Morelli SA, Sacchet MD, & Zaki J (2015). Common and distinct neural correlates of personal and vicarious reward: A quantitative meta-analysis. NeuroImage, 112(Ann. N. Y. Acad. Sci. 1191 2010), 244–253. 10.1016/j.neuroimage.2014.12.056 [PubMed: 25554428]

Nicolle A, Klein-Flügge MC, Hunt LT, Vlaev I, Dolan RJ, & Behrens T (2012). An Agent Independent Axis for Executed and Modeled Choice in Medial Prefrontal Cortex. Neuron, 75(6), 1114–1121. 10.1016/j.neuron.2012.07.023 [PubMed: 22998878]

O'Doherty J, Dayan P, Schultz J, Deichmann R, Friston K, & Dolan RJ (2004). Dissociable Roles of Ventral and Dorsal Striatum in Instrumental Conditioning. Science, 304(5669), 452–454. 10.1126/science.1094285 [PubMed: 15087550]

O'DOHERTY J, HAMPTON A, & KIM H (2007). Model-Based fMRI and Its Application to Reward Learning and Decision Making. Annals of the New York Academy of Sciences, 1104(1), 35–53. 10.1196/annals.1390.022 [PubMed: 17416921]

O'Doherty J, Winston J, Critchley H, Perrett D, Burt D ., & Dolan R . (2003). Beauty in a smile: the role of medial orbitofrontal cortex in facial attractiveness. Neuropsychologia, 41(2), 147–155. 10.1016/s0028-3932(02)00145-8 [PubMed: 12459213]

Ogawa A, Ueshima A, Inukai K, & Kameda T (2018). Deciding for others as a neutral party recruits risk-neutral perspective-taking: Model-based behavioral and fMRI experiments. Scientific Reports, 8(1), 12857 10.1038/s41598-018-31308-6 [PubMed: 30150657]

Olsson A, Nearing KI, & Phelps EA (2007). Learning fears by observing others: the neural systems of social fear transmission. Social Cognitive and Affective Neuroscience, 2(1), 3–11. 10.1093/scan/nsm005 [PubMed: 18985115]

Olsson A, & Phelps EA (2007). Social learning of fear. Nature Neuroscience, 10(9), 1095–1102. 10.1038/nn1968 [PubMed: 17726475]

Oosterhof NN, & Todorov A (2008). The functional basis of face evaluation. Proceedings of the National Academy of Sciences of the United States of America, 105(32), 11087–11092. 10.1073/pnas.0805664105 [PubMed: 18685089]

Rangel A, Camerer C, & Montague RP (2008). A framework for studying the neurobiology of value-based decision making. Nature Reviews Neuroscience, 9(7), 545–556. 10.1038/nrn2357 [PubMed: 18545266]

Ratcliff R, & McKoon G (2008). The Diffusion Decision Model: Theory and Data for Two-Choice Decision Tasks. Neural Computation, 20(4), 873–922. 10.1162/neco.2008.12-06-420 [PubMed: 18085991]

Ray D, King-casas B, Montague RP, & Dayan P (2008). Bayesian Model of Behaviour in Economic Games.

Roskies AL (1999). The Binding Problem. Neuron, 24(1), 7–9. 10.1016/s0896-6273(00)80817-x [PubMed: 10677022]

Ruff CC, & Fehr E (2014). The neurobiology of rewards and values in social decision making. Nature Reviews Neuroscience, 15(8), nrn3776 10.1038/nrn3776

Rutledge RB, Dean M, Caplin A, & Glimcher PW (2010). Testing the Reward Prediction Error Hypothesis with an Axiomatic Model. The Journal of Neuroscience, 30(40), 13525–13536. 10.1523/jneurosci.1747-10.2010 [PubMed: 20926678]

Sanfey AG, Rilling JK, Aronson JA, Nystrom LE, & Cohen JD (2003). The Neural Basis of Economic Decision-Making in the Ultimatum Game. Science, 300(5626), 1755–1758. 10.1126/science.1082976 [PubMed: 12805551]

Schultz J, & Bülthoff HH (2019). Perceiving animacy purely from visual motion cues involves intraparietal sulcus. NeuroImage, 197, 120–132. 10.1016/j.neuroimage.2019.04.058 [PubMed: 31028922]

Schultz W, Dayan P, & Montague RP (1997). A Neural Substrate of Prediction and Reward. Science, 275(5306), 1593–1599. 10.1126/science.275.5306.1593 [PubMed: 9054347]

Smith DV, Clithero JA, Boltuck SE, & Huettel SA (2014). Functional connectivity with ventromedial prefrontal cortex reflects subjective value for social rewards. Social Cognitive and Affective Neuroscience, 9(12), 2017–2025. 10.1093/scan/nsu005 [PubMed: 24493836]

Stanley DA (2016). Getting to know you: general and specific neural computations for learning about people. Social Cognitive and Affective Neuroscience, 11(4), 525–536. 10.1093/scan/nsv145 [PubMed: 26656563]

Strombach T, Weber B, Hangebrauk Z, Kenning P, Karipidis II, Tobler PN, & Kalenscher T (2015). Social discounting involves modulation of neural value signals by temporoparietal junction. Proceedings of the National Academy of Sciences, 112(5), 1619–1624. 10.1073/pnas.1414715112

Sul S, Tobler PN, Hein G, Leiberg S, Jung D, Fehr E, & Kim H (2015). Spatial gradient in value representation along the medial prefrontal cortex reflects individual differences in prosociality. Proceedings of the National Academy of Sciences, 112(25), 7851–7856. 10.1073/pnas.1423895112

Suzuki S, Adachi R, Dunne S, Bossaerts P, & O'Doherty JP (2015). Neural Mechanisms Underlying Human Consensus Decision-Making. Neuron, 86(2), 591–602. 10.1016/j.neuron.2015.03.019 [PubMed: 25864634]

Suzuki S, Cross L, & O'Doherty JP (2017). Elucidating the underlying components of food valuation in the human orbitofrontal cortex. Nature Neuroscience, 20(12), 1780–1786. 10.1038/s41593-017-0008-x [PubMed: 29184201]

Suzuki S, Harasawa N, Ueno K, Gardner JL, Ichinohe N, Haruno M, Cheng K, & Nakahara H (2012). Learning to Simulate Others' Decisions. Neuron, 74(6), 1125–1137. 10.1016/j.neuron.2012.04.030 [PubMed: 22726841]

Takahashi H, Kato M, Matsuura M, Mobbs D, Suhara T, & Okubo Y (2009). When Your Gain Is My Pain and Your Pain Is My Gain: Neural Correlates of Envy and Schadenfreude. Science, 323(5916), 937–939. 10.1126/science.1165604 [PubMed: 19213918]

Todorov A, Baron SG, & Oosterhof NN (2008). Evaluating face trustworthiness: a model based approach. Social Cognitive and Affective Neuroscience, 3(2), 119–127. 10.1093/scan/nsn009 [PubMed: 19015102]

Todorov A, Mandisodza AN, Goren A, & Hall CC (2005). Inferences of Competence from Faces Predict Election Outcomes. Science, 308(5728), 1623–1626. 10.1126/science.1110589 [PubMed: 15947187]

van den Bos W, Talwar A, & McClure SM (2013). Neural Correlates of Reinforcement Learning and Social Preferences in Competitive Bidding. The Journal of Neuroscience, 33(5), 2137–2146. 10.1523/jneurosci.3095-12.2013 [PubMed: 23365249]

Wegrzyn M, Riehle M, Labudda K, Woermann F, Baumgartner F, Pollmann S, Bien CG, & Kissler J (2015). Investigating the brain basis of facial expression perception using multi-voxel pattern analysis. Cortex, 69, 131–140. 10.1016/j.cortex.2015.05.003 [PubMed: 26046623]

Will G-J, Rutledge RB, Moutoussis M, & Dolan RJ (2017). Neural and computational processes underlying dynamic changes in self-esteem. ELife, 6, e28098 10.7554/elife.28098 [PubMed: 29061228]

Winston JS, Strange, O'Doherty J, & Dolan RJ (2002). Automatic and intentional brain responses during evaluation of trustworthiness of faces. Nature Neuroscience, 5(3), 277–283. 10.1038/nn816 [PubMed: 11850635]

Wittmann MK, Kolling N, Faber NS, Scholl J, Nelissen N, & Rushworth M (2016). Self-Other Mergence in the Frontal Cortex during Cooperation and Competition. Neuron, 91(2), 482–493. 10.1016/j.neuron.2016.06.022 [PubMed: 27477020]

Wittmann MK, Lockwood PL, & Rushworth MF (2018). Neural Mechanisms of Social Cognition in Primates. Annual Review of Neuroscience, 41(1), 99–118. 10.1146/annurev-neuro-080317-061450

Xiang T, Ray D, Lohrenz T, Dayan P, & Montague RP (2012). Computational Phenotyping of Two-Person Interactions Reveals Differential Neural Response to Depth-of-Thought. PLoS Computational Biology, 8(12), e1002841 10.1371/journal.pcbi.1002841 [PubMed: 23300423]

Yoshida W, Dolan RJ, & Friston KJ (2008). Game Theory of Mind. PLoS Computational Biology, 4(12), e1000254 10.1371/journal.pcbi.1000254 [PubMed: 19112488]

Yoshida W, Seymour B, Friston KJ, & Dolan RJ (2010). Neural Mechanisms of Belief Inference during Cooperative Games. The Journal of Neuroscience, 30(32), 10744–10751. 10.1523/jneurosci.5895-09.2010 [PubMed: 20702705]

Zentall TR (2012). Perspectives on observational learning in animals. Journal of Comparative Psychology, 126(2), 114 10.1037/a0025381 [PubMed: 21895354]

Zhu L, Mathewson KE, & Hsu M (2012). Dissociable neural representations of reinforcement and belief prediction errors underlie strategic learning. Proceedings of the National Academy of Sciences, 109(5), 1419–1424. 10.1073/pnas.1116783109

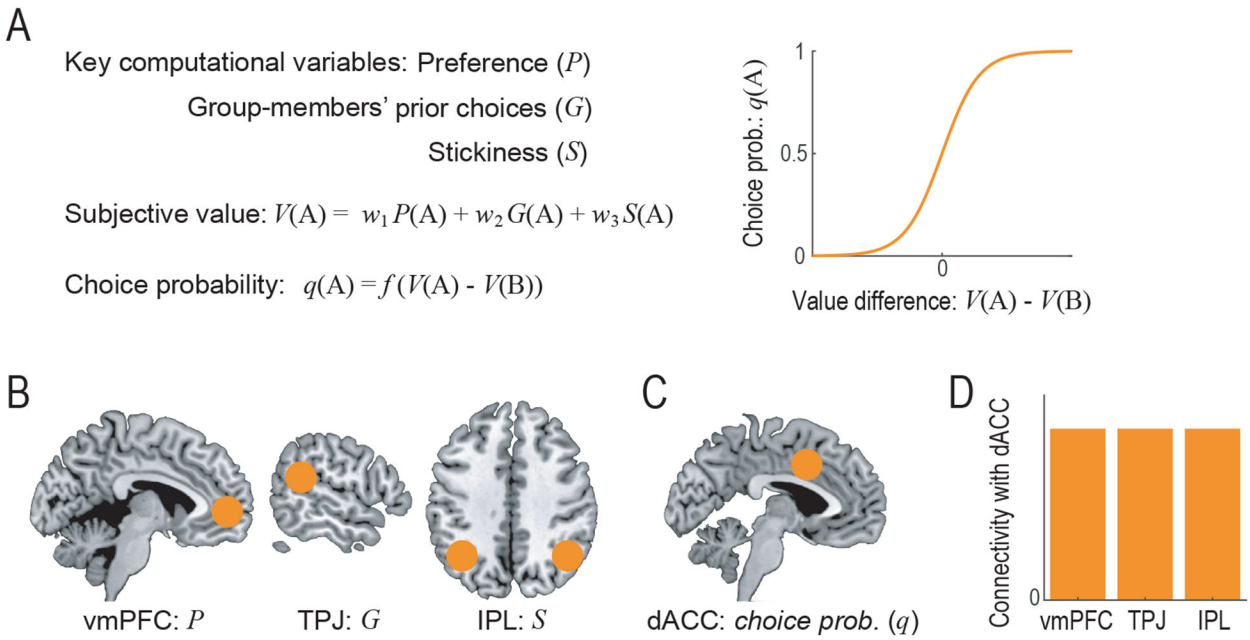

Figure 1 -

Neural mechanism underlying human consensus decision-making (Suzuki et al., 2015).

(A) Illustration of the computational model. Subjective value of each option is computed through integrating one's own preference, group-members' prior choices and one's inference about how much each option is stuck to by the other group-members, and then finally converted to the choice probability.

(B) Neural correlates of the three key computational variables. vmPFC: ventromedial prefrontal cortex; TPJ: temporoparietal junction; and IPL: inferior parietal lobule.

(C) Neural correlates of the integrated choice probability. dACC, dorsal anterior cingulate cortex.

(D) Functional connectivity between dACC and each of the three regions individually tracking the key computational variables.

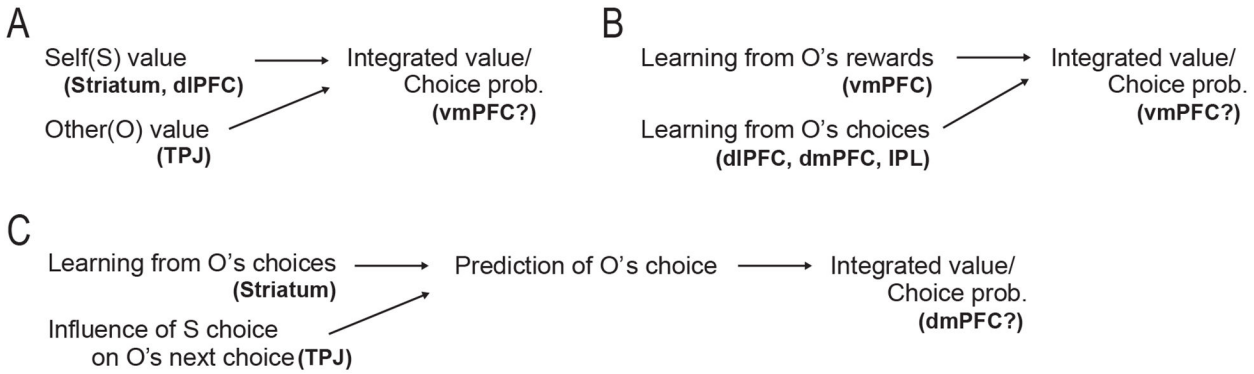

Figure 2 -

Schematic illustrations of example computational models for social value-based decision-making and their neural correlates.

(A) Decision-making for others (Fukuda et al., 2019; Hutcherson et al., 2015). Overall value and choice probability of each option is computed by integrating information about self and other reward values. S: self; O: other; dlPFC: dorsolateral prefrontal cortex; TPJ: temporoparietal junction; and vmPFC: ventromedial prefrontal cortex.

(B) Decision-making through observing others (Burke et al., 2010; Suzuki et al., 2012). Value and choice probability of each option is computed by integrating two types of learning from others' rewards and choices. dmPFC: dorsomedial prefrontal cortex; and IPL: inferior parietal lobule.

(C) Decision-making in strategic interactions (Hampton et al., 2008; Hill et al., 2017). Value of each option is computed by the prediction of the other's choice that integrates learning from the other's past choices and higher-order inference about influence of self-choice on the other's next choice.