David V. Smith · Patricia L. Lockwood ·
Dominic S. Fareri *Editors*

Neuroeconomics: Core Topics and Current Directions

Springer

Neuroeconomics: Core Topics and Current Directions

David V. Smith • Patricia L. Lockwood •
Dominic S. Fareri
Editors

Neuroeconomics: Core Topics and Current Directions

Springer

Foundations of Social Decision-Making

Cătălina Elena Rățală and Alan Gerard Sanfey

Abstract This chapter provides an overview of some of the foundational work and methods on neuroeconomic approaches to studying social decision-making, that is, where individuals make choices in social contexts, balancing personal preferences with the values, beliefs, and expectations of others. Outlining three core aspects of social decision-making—trust, cooperation, and resource allocation—we discuss the experimental paradigms used to examine each, and highlight the insights gained into the psychological and neural mechanisms underlying social interactions. We conclude with a discussion of new approaches that have the potential to expand our current understanding of social decision-making.

1 Introduction and Overview

Imagine you are in a restaurant, choosing what to eat for lunch. Your choice will depend on your own personal preferences, constraints, values, and so on. But what if you were having lunch with a close friend, who asked you to order on their behalf? Or what if you were having a work lunch and were concerned about how you will be perceived by your dining partners? These latter examples illustrate situations when we must additionally incorporate social, other-regarding, preferences into our decision calculus. In fact, many of the important choices we face in life, from political choices to ethical judgments, happen within a social context or involve a social component. Here, we use the term *social decision-making* to refer specifically to instances where one needs to consider not only one's own values, beliefs, or expectations, but also those of others who may be impacted by the outcome of the decision (Rilling & Sanfey, 2011).

C. E. Rățală (✉)
Donders Institute (Radboud University), Nijmegen, Netherlands
e-mail: catalina.ratala@donders.ru.nl

A. G. Sanfey
Donders Institute and Behavioural Science Institute (Radboud University), Nijmegen, Netherlands

D. V. Smith et al. (eds.), *Neuroeconomics: Core Topics and Current Directions*,
https://doi.org/10.1007/978-3-032-02925-6_5

Social decision scenarios are often characterized by high complexity and uncertainty, sometimes providing us with incomplete or even contradictory pieces of information. Understanding how we decide in these contexts, and what factors play a role in these choices, is important not only on an individual level but also for society as a whole, as this can provide vital information for improving policy prescriptions across many domains (Stallen & Sanfey, 2015). By using methodology from different fields, primarily Psychology, Economics, and Neuroscience, the discipline of Neuroeconomics can offer useful insights into why, and how, we make our social decisions (Camerer, 2007; Sanfey et al., 2006).

The aim of the current chapter is to outline some of the foundational methodologies and findings across Neuroeconomics which capture the complexity of social choice. To do this, we will focus on three core concepts pertaining to choice in social contexts—trust, cooperation, and resource allocation—and describe both the experimental paradigms typically employed to address these issues, as well as some fundamental findings Neuroeconomics has uncovered to date.

2 Trust and Reciprocity

Trust refers to the expectation that another party will behave in a reliable manner in situations of social and economic exchange, especially situations when the other can abuse this trust for their own self-interest. For example, when you lend money to a colleague, you act in trust, expecting the colleague will repay their debt. Reciprocity extends trust to also involve the possibility of mutual exchange, where your initial act of trust is acknowledged and returned in a manner that demonstrates reliability. Both concepts are crucial in reinforcing initial trust and facilitating subsequent cooperation.

In order to study the dynamics of trust and reciprocity, researchers have typically used paradigms inspired by the mathematical framework of Game Theory (von Neumann & Morgenstern, 1944). One such experimental paradigm is the Trust Game (Berg et al., 1995) (see Fig. 1a for details). The amount of money the Investor offers quantifies the trust that has been placed in the Trustee: higher investments indicate higher trust. Similarly, the money returned by the Trustee represents their level of reciprocity. The decision to initially trust a partner is influenced by various cognitive-affective, social, and contextual factors, and is a function of the information available to the Investor as well as their ability to infer the trustworthiness (i.e., the future likelihood of reciprocation) of their interaction partner, taking into account the potential risks of trusting another (Balliet & Van Lange, 2013; Hancock et al., 2023). Standard experimental findings using the Trust Game illustrate a reasonably high level of trust in anonymous partners, shown by people's willingness to invest about half of their initial endowment on average (Berg et al., 1995). Trustees in turn typically return slightly less than half of the amount they receive, illustrating a general preference for at least some degree of reciprocity. These findings are quite robust (see Johnson & Mislin, 2011 for a meta-analysis) and

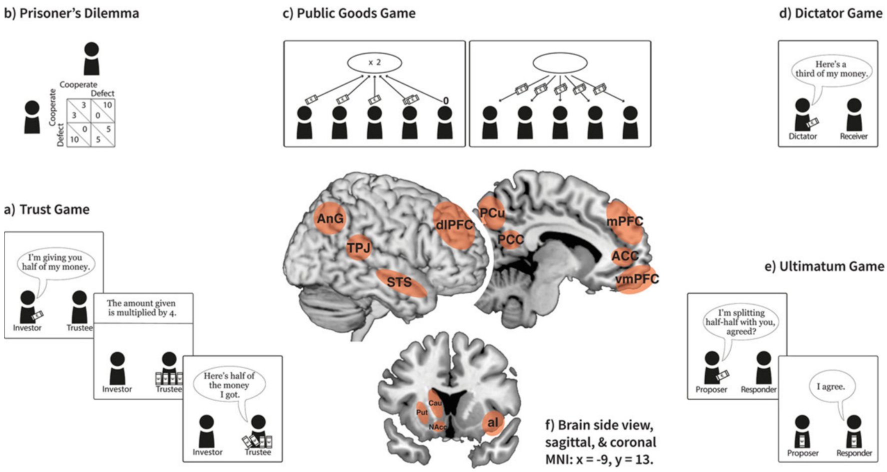

Fig. 1 Overview of classical experimental paradigms (**a**) **Trust Game**: This is a two-player game, where the Investor can send an amount of their money to the Trustee. The Trustee receives the quadrupled amount, and they can choose to reciprocate or not. The game isolates trust and reciprocity: the Investor must trust the Trustee to return a fair share, and the Trustee must decide whether or not to honor that trust. (**b**) **Prisoner's Dilemma.** This is a two-player game where each has two options: cooperate or defect. If both players cooperate, they both receive a moderate reward. If one player defects while the other cooperates, the defector gets a large reward and the cooperator gets no reward. The dilemma arises because the incentive to defect for personal gain can lead both to choose defection, resulting in a bad outcome for both. (**c**) **Public Goods Game.** It involves multiple players, who each receive an initial endowment. Players then decide how much to contribute to a public pot, keeping the rest for themselves. Contributions are multiplied by a specified factor, then shared *equally* among all. The dilemma is that optimal collective gains come from full contribution by all, but free-riders (who contribute nothing) can benefit from the public good while keeping their own initial amount. The game examines cooperative behavior and how individuals balance personal versus collective gain. (**d**) **Dictator Game.** This has two players. The Dictator is provided with an endowment and decides how much to share with the Recipient, retaining the remainder. The Recipient has no say in the matter. The game explores "pure" altruism, as the Dictator is under no obligation to share, but may choose to do so based on personal values or social norms. (**e**) **Ultimatum Game.** In this two-player game, the Proposer is given an endowment and proposes to the Responder how to divide this money. The Responder then must either accept or reject the offer. If the Responder accepts, the money is split as proposed. If the Responder rejects, neither player receives *anything*. In either event, the game is over. The game tests fairness and bargaining behavior, as the Proposer must make an offer that the Responder finds acceptable, balancing personal gain with the risk of rejection. (**f**) **Overview of brain areas associated with social decision-making**. In the side view: dorsal lateral prefrontal cortex (dmPFC), angular gyrus (AnG), temporal parietal junction (TPJ), and superior temporal sulcus (STS). In the sagittal view: medial prefrontal cortex (mPFC), anterior cingulate cortex (ACC), ventromedial prefrontal cortex (vmPFC), posterior cingulate cortex (PCC), and precuneus (PCu). In the coronal view: anterior insula (aI), putamen (Put), caudate (Cau), and nucleus accumbens (NAcc). The striatum comprises the putamen, caudate, and NACC. Insula and striatum are bilateral; display is unilateral for visual purposes

make clear that, contrary to the standard economic predictions of pure self-interest and hence zero investments and zero returns, players generally demonstrate both trusting and reciprocal behavior (Camerer, 2003).

These decisions to trust and to reciprocate are grounded in both implicit social evaluations (facial features, in-group considerations, biases, personality traits, and so on), as well as explicit social learning driven by previous interactions or overt social information (Mende-Siedlecki et al., 2013; see Fareri, 2019 for a mini-review). With regard to implicit factors, people often estimate how much they trust a complete stranger in just a glance, relying on extremely rapid mechanisms (at the level of milliseconds) of facial evaluation (Delgado et al., 2005; Oosterhof & Todorov, 2008; Todorov et al., 2009; Willis & Todorov, 2006). Though it should be noted that these rapid estimations are typically not very accurate, nonetheless neuroimaging evidence suggests these processes have their neural substrates in the amygdala (Engell et al., 2007), which is responsive to untrustworthy faces, as well as the medial prefrontal cortex (mPFC) and the precuneus (Todorov & Engell, 2008), which conversely show greater neural activity for (moderately) trustworthy faces. Crucially, when an Investor observes reciprocal behavior in their interaction partner, activation in the striatum, part of the brain's reward system, has been consistently recorded (Rilling et al., 2002), further reinforcing the initial trust decision.

Studies that focus on more explicit signals of trust and reciprocity often employ repeated-round Trust Games, where Investors have the opportunity to see exactly how their partners, the Trustees, respond to their trust over time, and vice versa. Factors such as prior trust experience or the level of trust bestowed by the Investor can substantially change the reciprocation response (Fareri, 2019). During the initial trust-building stages, the mPFC is involved in assessing trustworthiness and predicting future behavior while integrating information about social context and past experiences (Krueger et al., 2007; Krueger & Grafman, 2008). This process is in line with the idea that past occurrences serve as reliable indicators of future behavior (Axelrod & Hamilton, 1981; King-Casas et al., 2005). Furthermore, trust can be well captured via computational models utilizing prediction error-driven updates (Chang et al., 2010; Doll et al., 2009; Fareri et al., 2015). FMRI results have shown that the reputation of others can be encoded in the striatum, with reciprocity eliciting stronger responses than betrayal (Delgado, 2007; see Bellucci et al., 2017 for a meta-analysis). Across several interactions, the neural signals from the striatum shift earlier in time to help in anticipating the trust behavior of the interaction partner (King-Casas et al., 2005). Overall, reciprocity acts as both a socially rewarding experience and a learning signal to guide future trust decisions (Phan et al., 2010; Rilling et al., 2002), with areas such as the mPFC, striatum, and ACC being involved in positive reciprocation, while the anterior Insula seems to track negative reciprocation (see Fig. 1f). Since competing psychological accounts have been proposed for what may drive reciprocation behavior, for example, inequity aversion (Fehr & Schmidt, 1999; Tricomi et al., 2010) and guilt aversion (Battigalli & Dufwenberg, 2007; Dufwenberg & Gneezy, 2000), computational modelling offers a powerful framework for addressing these hypotheses. Chang and colleagues (2011) used a utility-based computational model (see Box 1) to illustrate how individual differences in beliefs and guilt sensitivity affect recruitment of the TPJ, insula, dACC, and dorsolateral PFC (see Fig. 1f). Further research using computational modelling and fMRI suggests the recruitment of several distinct strategies in accordance with those suggested above, as well as the ability to flexibly switch between these strategies in order to serve the twin goals of maintaining reputation while also maximizing one's own return (van Baar et al., 2019).

Box 1: Methodology

Computational Modeling. Computational modeling provides a quantifiable framework for studying individual choice in complex social environments. Mathematical models, often inspired from economics, are used to describe specific decision-related processes, such as learning, valuation, and social preferences. Different types of models can help capture links between psychological constructs, behavioral measures, and neural measurements. For instance, reinforcement learning models capture how individuals adjust their behavior over time based on rewards and punishments, shedding light on learning mechanisms in decision-making. Utility models quantify how individuals trade off social and monetary preferences in prosocial choices by assigning subjective values to different outcomes. Drift diffusion models describe the process of evidence accumulation, where information is integrated until a decision threshold is reached. Meanwhile, Bayesian models characterize how individuals update their beliefs about others' intentions by probabilistically incorporating new information as a function of prior knowledge or beliefs.

Physiological Measurements. Physiological responses provide valuable insights into the emotional and autonomic processes underlying decision-making. **Skin conductance response (SCR)** measures phasic changes in skin conductance driven by the sympathetic nervous system activity, reflecting emotional arousal. As a sensitive index of physiological arousal, SCR is widely used to study emotional engagement in decision-making and social interactions. **Heart rate variability (HRV)** indicates the balance between sympathetic and parasympathetic nervous system activity, which plays a role in stress regulation and decision-making under uncertainty. These physiological signals, when combined with computational models and/or neuroimaging, can offer a broader understanding of how affective and physiological states affect our choices. Techniques such as **eye-tracking** provide real-time insights into attention allocation, cognitive effort, and information processing. Eye-tracking measures fixation duration, saccades, and gaze patterns, revealing how individuals visually explore choice options and social cues.

Pharmacological Interventions. Pharmacological interventions involve the administration of drugs (e.g., methylphenidate, L-Dopa, atomoxetine) that

(continued)

modulate various neurotransmitters to examine their role in behavior and brain function. By comparing drug-induced changes across different populations, this approach provides insights into the neurochemical mechanisms underlying cognition and potential treatments for clinical conditions.

Electroencephalography (EEG): EEG is also a popular method in neuroeconomics, and it measures the electrical activity in the brain as a function of voltage changes in neurons. This method has excellent temporal resolution (on the millisecond scale), but poor spatial resolution. EEG can address questions related to tracking the time course of various neural processes, especially fast-changing ones like attention.

Magnetoencephalography (MEG): MEG measures the magnetic fields generated by neuronal electrical activity. Similar to EEG, MEG has excellent temporal resolution, but also better spatial resolution relative to EEG. Researchers will sometimes combine MEG and fMRI, to benefit from the complementary spatial and temporal information.

Functional Magnetic Resonance Imaging (fMRI). FMRI is one of the widely used methods in neuroeconomics. It measures metabolic changes, thus providing an indirect way to correlate mental processes with neural substrates. While it has good spatial resolution, its temporal resolution is limited. Various statistical approaches are used for analyzing fMRI data, focusing either on univariate average signal (univariate) or a distributed pattern (multivariate), making this method ideal for answering questions pertaining to where in the brain changes occur as a function of experimental task.

Positron Emission Tomography (PET). PET uses radioactive tracers to directly measure neurotransmitter function, receptor availability, and brain metabolism. It is essential in studying dopamine, serotonin, and opioid systems, which regulate reward processing.

Transcranial Magnetic Stimulation (TMS). TMS is a noninvasive brain stimulation technique that uses magnetic fields to induce electrical currents in targeted brain areas. This approach allows researchers to temporarily disrupt (or enhance) neural activity, making it a powerful tool for establishing causal relationships between brain regions and cognitive functions. TMS is frequently used in combination with neuroimaging to test specific hypotheses about brain function.

Transcranial Direct Current Stimulation (tDCS) and Transcranial Alternating Current Stimulation (tACS). Both tDCS and tACS are noninvasive neuromodulatory techniques that deliver low-intensity electrical currents to modulate neural activity. tDCS applies a constant direct current to alter cortical excitability, while tACS uses an oscillatory current at specific frequencies to influence brain rhythms. These techniques are often paired with behavioral tasks and neuroimaging (EEG, fMRI) to examine how targeted neural modulation affects decision-making, learning, and social cognition.

(continued)

Like TMS, they provide causal evidence for the role of specific brain regions in cognitive processes, making them valuable tools in experimental neuroeconomics.

Pharmacologically, a key neurotransmitter related to trust, trust learning, and reciprocity is dopamine. Primarily associated with reward receipt, reward anticipation, and prediction errors, several studies have shown that dopamine pathways become active during trust interactions, particularly in the striatum (Schultz, 2002, 2013; Izuma et al., 2008). Moreover, the evaluation of the trustworthiness of others has also been linked to the dopaminergic system (Krueger et al., 2007). In addition to dopamine, experimental work also connects other neurotransmitters, such as serotonin (Crockett et al., 2008, 2010), or hormones, such as testosterone and oxytocin, to trust behavior. However, findings to date are mixed, suggesting complex patterns of interaction are at play.

3 Cooperation, Coordination, and Competition

Cooperation refers to the joint investment of resources in order to achieve a mutually beneficial outcome. It is considered one of the cornerstones of human civilization (Tomasello & Vaish, 2013), and relies heavily on the ability to engage in mutual exchanges, which of course are potentially risky and not necessarily immediately or directly beneficial for oneself (Fehr & Fischbacher, 2003; see Stallen & Sanfey, 2013 for a review).

For instance, imagine you are working with a colleague on a project; each of you might have different strengths and weaknesses, and so collaboration can improve the resulting outcome relative to what could have been achieved by working alone. However, you might also want to guard against being exploited by your colleague, for example, if you do the lion's share of the work, but they take the credit. So, what are the factors that can facilitate or hinder cooperation? Why do people cooperate (or not), and how can Neuroeconomics help answer such questions?

To investigate cooperation experimentally, researchers have primarily used two classic paradigms or "coordination games": the Prisoner's Dilemma (Davis & Holt, 1993; Sally, 1995) and the Public Goods Game (Ledyard, 1995; see Fig. 1b, c for details). The Prisoner's Dilemma examines cooperative behavior as a function of joint outcomes. Findings here indicate that people prefer smaller mutual rewards gained via cooperation to higher personal rewards obtained by betraying the other (Davis & Holt, 1993), again contradicting theoretical predictions of pure self-interest (Sally, 1995). However, potential financial gains do matter, and both the increase of rewards or the introduction of punishment in this task has a moderate positive effect on cooperation (see Balliet et al., 2011 for a meta-analysis). Some of the primary factors that have been shown to impact cooperation in the Prisoner's Dilemma pertain to the payoff structure (Pruitt, 1967; Mengel, 2018; Gächter et al., 2021) and to how social values are weighted (see Bogaert et al., 2008 for a review). For example, if individuals find it more rewarding to behave prosocially (i.e., they value the benefit of others over their own financial gain), then they will cooperate more often. Additionally, when people interact together more than once, strategic considerations of reputation and reciprocity are observed. One of the most prominent and successful strategies in repeated interactions is the so-called "tit-for-tat," where each individual mirrors in the current round of the game the strategy their opponent chose in the previous round (Axelrod & Hamilton, 1981).

Experimental studies have explored the neural basis of cooperative behavior in the Prisoner's Dilemma (Rilling et al., 2008; Suzuki et al., 2011; Thompson et al., 2021; see Rilling & Sanfey, 2011 for a review). Accumulated evidence has identified the PFC, the ACC, the striatum, and the superior temporal sulcus (STS) as key brain regions involved in deciding whether to cooperate or defect. Cooperation in repeated games has been linked to activity in reward-related areas such as the striatum, ventromedial PFC, and orbitofrontal cortex (Haber & Knutson, 2010; Rilling et al., 2002). In addition, the posterior STS has been identified as crucial for mentalizing about the feelings, intentions, and strategies of others (Hampton et al., 2008; see Hein & Knight, 2008 for a review). Empathy has been linked with prosocial behavior (see Lockwood et al., 2016, for a review) and decisions to cooperate, involving areas such as insula, the ACC (Apps et al., 2016; Singer et al., 2004, 2006; see Lockwood & Wittmann, 2018 for a review), and the PFC (Masten et al., 2011).

Expanding on cooperative dynamics beyond two-player interactions, the Public Goods Game highlights social norms and collective action (Fehr & Gächter, 2000). While behavior here emphasizes the beneficial power of cooperation for the group (Stallen & Sanfey, 2015; Vives & FeldmanHall, 2018), decision-making is also susceptible to strategic considerations by individuals who prioritize personal gain over group welfare, often termed "free-riding" (Isaac & Walker, 1988). Experimental findings show collective investment behavior, with individuals on average contributing slightly less than half of their endowment to the common "pot" (Ledyard, 1995). This suggests motivations that go beyond purely financial incentives, and illustrates a willingness to contribute based on prosocial considerations (Fehr & Gächter, 2002; Rilling et al., 2002). A meta-analysis by Pletzer et al. (2018) has shown that this decision to cooperate is partly driven by the expectations about others' cooperative behavior. Simply making choices which improve the well-being of others recruits brain regions previously associated with socio-cognitive and empathic processes (see Bellucci et al., 2020 for a meta-analysis), such as the ventromedial PFC, the dorsolateral PFC, and the dorsal posterior cingulate cortex (see Fig. 1f). Additionally, accumulating evidence suggests that during various social interactions which involve mimicry or cooperation, there is also significant synchronization of brain activity in areas related to social cognition, for example, theory of mind or goal-oriented behavior (Dumas et al., 2010). Though evidence indicates that cooperative preferences are rather stable and generalized across various contexts (Peysakhovich et al., 2014), recent studies show that even when decision outcomes are similar, the brain mechanisms that underlie this behavior might differ (Rhoads et al., 2021; van Baar et al., 2019).

4 Resource Allocation

A final category of interactive scenarios that are important for understanding social choice involves decisions about the allocation and distribution of resources. Understanding how people trade off various competing motivations in these contexts offers crucial insights into the neural and psychological processes that underpin human social behavior (Rilling & Sanfey, 2011).

Imagine a scenario where an individual decides to donate a part of their salary to support the search for a disease treatment. This is considered altruistic behavior, as the donor themselves is unlikely to directly benefit from this. Among the several psychological factors that impact resource allocation, fairness and social preferences are some of the most commonly studied. The experimental paradigms used to investigate fairness employ tasks which attempt to tease apart how people balance social and monetary reward.

The Dictator Game (Kahneman et al., 1986) stands out as one of the most straightforward yet revealing paradigms in use (see Fig. 1d). Despite this simplicity, the Dictator Game has yielded substantial insights into human behavior. Contrary to the classical predictions of pure self-interest, most Dictators choose to give nonzero amounts to their game partners (Camerer, 1997, 2003; see Engel, 2011 for a meta-analysis), with a 50–50 split being the most frequent allocation (Camerer & Thaler, 1995; Fehr & Gächter, 2002). Although at first glance this finding seems to reveal an intrinsic preference for fairness that extends well beyond self-interest, literature suggests there are several other complex motivations at play (Andreoni & Bernheim, 2009; Murnighan et al., 2001).

One strand of research suggests that Dictators choose to share a part of their endowment due to a strong aversion to inequity or inequality (Fehr & Schmidt, 1999; Fehr et al., 2006). Krupka and Weber (2013) provide some evidence for these social preferences, showing that individuals exhibit a stable willingness to forfeit monetary gains to engender socially appropriate outcomes. Similarly, Schulz et al. (2014) illustrate that under high cognitive load (meant to decrease deliberation) participants choose fair allocations more often than under low cognitive load. This suggests fairness may well be the "default" option in the Dictator Game, with selfish behavior being the result of deliberative thinking. However, there appears to be substantial heterogeneity in allocation choices in this task. Meta-analytic findings indicate that fairness motivations are also moderated by socioeconomic factors and societal values: the more economically developed or individualistic a society, the less Dictators are willing to share (Engel, 2011). Furthermore, in the context of this paradigm, a significant number of Dictators tend to exploit any potential exculpatory mechanism available, such as ambiguity, plausible deniability, or willful ignorance, in order to appear fair without actually paying the monetary cost required by fairness

(Dana et al., 2007; Haisley & Weber, 2010). Thus, fairness seems not only subject to framing, contextual factors, and situational constraints (Andreoni & Rao, 2011; Camerer, 2003; Eckel & Grossman, 1996; Grossman, 2014), but also to individual tendencies for opportunism (Barron et al., 2022; van Baar et al., 2019).

The Ultimatum Game (Güth et al., 1982) is another experimental paradigm particularly useful in the investigation of fairness, focusing on the strategic element of distributive choice (see Fig. 1e). Relative to the Dictator Game, Ultimatum Game offers tend to be greater, likely due to the second-order preferences, beliefs, and strategic considerations of the allocator (Chen et al., 2017; Wells & Rand, 2013; Ruff et al., 2013; Sazhin et al., 2024; Spitzer et al., 2007). That is, concerns that a low offer will be immediately rejected by the Responder, thus costing the Proposer any gain. Similarly to the Dictator game, the Ultimatum Game has spawned numerous variations, each shedding light on different aspects of fairness considerations. For example, multi-round versions with changing roles (Nowak et al., 2000) have revealed how reputational concerns influence fairness behavior. The paradigm has been instrumental in showing also that people are willing to sacrifice personal gains in order to punish unfair behavior, a phenomenon that has been termed "altruistic punishment" (Fehr & Gächter, 2002). Importantly, varying the stakes (Cameron, 1999) illustrates that although people tend to increase their offers as stakes grow larger, the willingness to reject unfair offers (typically defined as lower than 20% of the endowed amount) persists even with substantial sums of money (see Larney et al., 2019 for a meta-analysis). In 2003, Sanfey and colleagues showed that the anterior insula is one of the key regions involved in the processing of unfair offers, with greater activation for unfair offers compared to fair offers predictive of the rejection of unfair offers. This finding suggests that the affective response to unfairness, potentially linked to feelings of disgust or anger, plays a crucial role in redistribution decisions (see Gabay et al., 2014 for a meta-analysis) and emphasizes the important role of emotions in social decision-making, challenging purely rational models of economic choice. Additionally, the dlPFC has been associated with cognitive control and the ability to override automatic responses in various contexts, including fairness decisions. More recent neuroimaging findings suggest that these strategic decisions activate the inferior frontal gyrus, and in particular, enhance its connectivity with the TPJ, whereas anterior insula-angular gyrus connectivity appears to modulate strategic behavior based on individual differences in reward (Sazhin et al., 2024) (see Fig. 1f). Several studies (van't Wout et al., 2005; Knoch et al., 2006) used transcranial magnetic stimulation (TMS) to disrupt dlPFC function and found that participants were more likely to accept unfair offers when this region was inhibited (see Box 1). Tricomi et al. (2010) illustrate how both advantageous and disadvantageous inequality can activate reward-related brain regions. Norm compliance likewise involves activity in cortical areas linked to reward processing. For instance, studies indicate that adherence to social norms under threat of punishment depends on the right lateral prefrontal cortex (Ruff et al., 2013). Experimentally increasing excitability in this region via anodal tDCS boosted sanction-induced norm compliance, whereas decreasing excitability dampens compliance. Complementary work demonstrated that rLPFC stimulation exerted stronger effects when social interactions are made explicit. Interestingly, studies have shown striatal activation not only when individuals receive rewards themselves, but also when they observe others receiving rewards, especially in fair contexts (Moll et al., 2006). This suggests a neural basis for the positive feelings associated with fair distributions and provides some support for the intrinsic value of equality.

5 Conclusion

Neuroeconomic approaches to social decision-making have provided valuable insights into the understanding of complex social behaviors, as well as the understanding of their neural underpinnings. By combining rigorous experimental paradigms with advanced neuroscientific techniques, researchers have started to unravel how the brain makes decisions about trust and reciprocity, cooperation, and resource allocation.

Future research in social decision-making will likely focus on integrating these findings into more comprehensive models of behavior. Additionally, the use of new approaches and technologies, such as wearable devices, large data sets, and virtual reality paradigms, to name but a few, all promise to add to the ecological validity of the current tasks and advance our current understanding of these complex phenomena.

Several directions of research are particularly promising, for instance the use of computational modelling where leveraging multimodal data and frameworks from machine learning and game theory can enhance prediction accuracy, theory development, and computational phenotyping across populations (Ahn & Busemeyer, 2016; Collins & Shenhav, 2022; Gillan & Whelan, 2017; Lee, 2013; Montague et al., 2012; see Brazil et al., 2018 for a review). Furthermore, the rapid developments in AI (see Harris, 2024 for a review) allow researchers to understand and predict social and economic behavior through the analysis of massive, highly complex data (see Aoujil et al., 2023 for a bibliographic analysis). Emerging research also highlights the gut-brain axis as a modulator of social decision-making, with gut microbiota influencing stress responses, emotional regulation, and reward processing.

Ultimately, the cumulative findings to date in social decision-making not only refine our understanding of the distinct motivations driving prosocial behavior (Schreuders et al., 2018) but also offer pathways for developing interventions for individuals with atypical social processing (Bellucci et al., 2020). Furthermore, this knowledge can be a powerful tool when applied to real-world settings, offering policymakers valuable insights in the design of evidence-based strategies aimed at promoting better, more sustainable decision-making and prosocial behavior in various societal contexts.

References

Ahn, W.-Y., & Busemeyer, J. R. (2016). Challenges and promises translating computational tools into clinical practice. *Current Opinion in Behavioral Sciences, 11*, 1–7. https://doi.org/10.1016/j.cobeha.2016.02.001

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

Andreoni, J., & Rao, J. M. (2011). The power of asking: How communication affects selfishness, empathy, and altruism. *Journal of Public Economics, 95*(7), 513–520. https://doi.org/10.1016/j.jpubeco.2010.12.008

Aoujil, Z., Hanine, M., Flores, E. S., Samad, M. A., & Ashraf, I. (2023). Artificial intelligence and behavioral economics: A bibliographic analysis of research field. *IEEE Access, 11*, 139367–139394. https://doi.org/10.1109/ACCESS.2023.3339778

Apps, M. A., Rushworth, M. F., & Chang, S. W. (2016). The anterior cingulate gyrus and social cognition: Tracking the motivation of others. *Neuron, 90*(4), 692–707. https://doi.org/10.1016/j.neuron.2016.04.018

Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science, 211*(4489), 1390–1396.

Balliet, D., & Van Lange, P. A. (2013). Trust, conflict, and cooperation: A meta-analysis. *Psychological Bulletin, 139*(5), 1090. https://doi.org/10.1037/a0030939

Balliet, D., Mulder, L. B., & Van Lange, P. A. M. (2011). Reward, punishment, and cooperation: A meta-analysis. *Psychological Bulletin, 137*(4), 594–615. https://doi.org/10.1037/a0023489

Barron, K., Stüber, R., & van Veldhuizen, R. (2022). *Motivated motive selection in the lying-dictator game* (CESifo Working Paper No. 9911). Available at SSRN. https://doi.org/10.2139/ssrn.4215030

Battigalli, P., & Dufwenberg, M. (2007). Guilt in games. *American Economic Review, 97*(2), 170–176. https://doi.org/10.1257/aer.97.2.170

Bellucci, G., Chernyak, S. V., Goodyear, K., Eickhoff, S. B., & Krueger, F. (2017). Neural signatures of trust in reciprocity: A coordinate-based meta-analysis. *Human Brain Mapping, 38*(3), 1233–1248. https://doi.org/10.1002/hbm.23451

Bellucci, G., Camilleri, J. A., Eickhoff, S. B., & Krueger, F. (2020). Neural signatures of prosocial behaviors. *Neuroscience and Biobehavioral Reviews, 118*, 186–195. https://doi.org/10.1016/j.neubiorev.2020.07.006

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior, 10*(1), 122–142. https://doi.org/10.1006/game.1995.1027

Bogaert, S., Boone, C., & Declerck, C. (2008). Social value orientation and cooperation in social dilemmas: A review and conceptual model. *British Journal of Social Psychology, 47*(3), 453–480. https://doi.org/10.1348/014466607X244970

Brazil, I. A., van Dongen, J. D., Maes, J. H., Mars, R. B., & Baskin-Sommers, A. R. (2018). Classification and treatment of antisocial individuals: From behavior to biocognition. *Neuroscience & Biobehavioral Reviews, 91*, 259–277. https://doi.org/10.1016/j.neubiorev.2016.10.010

Camerer, C. F. (1997). Progress in behavioral game theory. *Journal of Economic Perspectives, 11*(4), 167–188. https://doi.org/10.1257/jep.11.4.167

Camerer, C. F. (2003). *Behavioral game theory: Experiments in strategic interaction* (p. xv, 550).

Camerer, C. F. (2007). Neuroeconomics: Using neuroscience to make economic predictions. *The Economic Journal, 117*(519), C26–C42. https://doi.org/10.1111/j.1468-0297.2007.02033.x

Camerer, C., & Thaler, R. H. (1995). Anomalies: Ultimatums, dictators and manners. *Journal of Economic Perspectives, 9*(2), 209–219. https://doi.org/10.1257/jep.9.2.209

Cameron, L. A. (1999). Raising the stakes in the ultimatum game: Experimental evidence from Indonesia. *Economic Inquiry, 37*(1), 47–59. https://doi.org/10.1111/j.1465-7295.1999.tb01415.x

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
Chang, L. J., Smith, A., Dufwenberg, M., & Sanfey, A. G. (2011). Triangulating the neural, psychological, and economic bases of guilt aversion. *Neuron, 70*(3), 560–572. https://doi.org/10.1016/j.neuron.2011.02.056
Chen, Y.-H., Chen, Y.-C., Kuo, W.-J., Kan, K., Yang, C. C., & Yen, N.-S. (2017). Strategic motives drive proposers to offer fairly in ultimatum games: An fMRI study. *Scientific Reports, 7*(1), 527. https://doi.org/10.1038/s41598-017-00608-8
Collins, A. G. E., & Shenhav, A. (2022). Advances in modeling learning and decision-making in neuroscience. *Neuropsychopharmacology, 47*(1), 104–118. https://doi.org/10.1038/s41386-021-01126-y
Crockett, M. J., Clark, L., Tabibnia, G., Lieberman, M. D., & Robbins, T. W. (2008). Serotonin modulates behavioral reactions to unfairness. *Science, 320*(5884), 1739–1739. https://doi.org/10.1126/science.1155577
Crockett, M. J., Clark, L., Hauser, M. D., & Robbins, T. W. (2010). Serotonin selectively influences moral judgment and behavior through effects on harm aversion. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 107*(40), 17433–17438. https://doi.org/10.1073/pnas.1009396107
Dana, J., Weber, R. A., & Kuang, J. X. (2007). Exploiting moral wiggle room: Experiments demonstrating an illusory preference for fairness. *Economic Theory, 33*(1), 67–80. https://doi.org/10.1007/s00199-006-0153-z
Davis, D., & Holt, C. (1993). *Experimental economics*. Princeton University Press. https://doi.org/10.1515/9780691233376
Delgado, M. R. (2007). Reward-related responses in the human striatum. *Annals of the New York Academy of Sciences, 1104*(1), 70–88. https://doi.org/10.1196/annals.1390.002
Delgado, M. R., Miller, M. M., Inati, S., & Phelps, E. A. (2005). An fMRI study of reward-related probability learning. *NeuroImage, 24*(3), 862–873. https://doi.org/10.1016/j.neuroimage.2004.10.002
Doll, B. B., Jacobs, W. J., Sanfey, A. G., & Frank, M. J. (2009). Instructional control of reinforcement learning: A behavioral and neurocomputational investigation. *Brain Research, 1299*, 74–94. https://doi.org/10.1016/j.brainres.2009.07.007
Dufwenberg, M., & Gneezy, U. (2000). Measuring beliefs in an experimental lost wallet game. *Games and Economic Behavior, 30*(2), 163–182. https://doi.org/10.1006/game.1999.0715
Dumas, G., Nadel, J., Soussignan, R., Martinerie, J., & Garnero, L. (2010). Inter-brain synchronization during social interaction. *PLoS One, 5*(8), e12166. https://doi.org/10.1371/journal.pone.0012166
Eckel, C. C., & Grossman, P. J. (1996). The relative price of fairness: Gender differences in a punishment game. *Journal of Economic Behavior & Organization, 30*(2), 143–158. https://doi.org/10.1016/S0167-2681(96)00854-2
Engel, C. (2011). Dictator games: A meta study. *Experimental Economics, 14*(4), 583–610. https://doi.org/10.1007/s10683-011-9283-7
Engell, A. D., Haxby, J. V., & Todorov, A. (2007). Implicit trustworthiness decisions: Automatic coding of face properties in the human amygdala. *Journal of Cognitive Neuroscience, 19*(9), 1508–1519. https://doi.org/10.1162/jocn.2007.19.9.1508
Fareri, D. S. (2019). Neurobehavioral mechanisms supporting trust and reciprocity. *Frontiers in Human Neuroscience, 13*. https://doi.org/10.3389/fnhum.2019.00271
Fareri, D. S., Chang, L. J., & Delgado, M. R. (2015). Computational substrates of social value in interpersonal collaboration. *The Journal of Neuroscience, 35*(21), 8170–8180. https://doi.org/10.1523/JNEUROSCI.4775-14.2015
Fehr, E., & Fischbacher, U. (2003). The nature of human altruism. *Nature, 425*(6960), 785–791. https://doi.org/10.1038/nature02043
Fehr, E., & Gächter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review, 90*(4), 980–994. https://doi.org/10.1257/aer.90.4.980

Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature, 415*(6868), 137–140. https://doi.org/10.1038/415137a

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

Fehr, E., Naef, M., & Schmidt, K. M. (2006). Inequality aversion, efficiency, and maximin preferences in simple distribution experiments: Comment. *American Economic Review, 96*(5), 1912–1917. https://doi.org/10.1257/aer.96.5.1912

Gabay, A. S., Radua, J., Kempton, M. J., & Mehta, M. A. (2014). The ultimatum game and the brain: A meta-analysis of neuroimaging studies. *Neuroscience & Biobehavioral Reviews, 47*, 549–558. https://doi.org/10.1016/j.neubiorev.2014.10.014

Gächter, S., Lee, K., Sefton, M., & Weber, T. O. (2021). *Risk, temptation, and efficiency in the one-shot prisoner's dilemma* (CESifo working paper no. 9449). Available at SSRN. https://doi.org/10.2139/ssrn.3979354

Gillan, C. M., & Whelan, R. (2017). What big data can do for treatment in psychiatry. *Current Opinion in Behavioral Sciences, 18*, 34–42. https://doi.org/10.1016/j.cobeha.2017.07.003

Grossman, Z. (2014). Strategic ignorance and the robustness of social preferences. *Management Science, 60*(11), 2659–2665. https://doi.org/10.1287/mnsc.2014.1989

Güth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior & Organization, 3*(4), 367–388. https://doi.org/10.1016/0167-2681(82)90011-7

Haber, S. N., & Knutson, B. (2010). The reward circuit: Linking primate anatomy and human imaging. *Neuropsychopharmacology, 35*(1), 4–26. https://doi.org/10.1038/npp.2009.129

Haisley, E. C., & Weber, R. A. (2010). Self-serving interpretations of ambiguity in other-regarding behavior. *Games and Economic Behavior, 68*(2), 614–625. https://doi.org/10.1016/j.geb.2009.08.002

Hampton, A. N., Bossaerts, P., & O'Doherty, J. P. (2008). Neural correlates of mentalizing-related computations during strategic interactions in humans. *Proceedings of the National Academy of Sciences, 105*(18), 6741–6746. https://doi.org/10.1073/pnas.0711099105

Hancock, P. A., Kessler, T. T., Kaplan, A. D., Stowers, K., Brill, J. C., Billings, D. R., Schaefer, K. E., & Szalma, J. L. (2023). How and why humans trust: A meta-analysis and elaborated model. *Frontiers in Psychology, 14*, 1081086. https://doi.org/10.3389/fpsyg.2023.1081086

Harris, L. T. (2024). The neuroscience of human and artificial intelligence presence. *Annual Review of Psychology, 75*(1), 433–466. https://doi.org/10.1146/annurev-psych-013123-123421

Hein, G., & Knight, R. T. (2008). Superior temporal sulcus—It's my area: Or is it? *Journal of Cognitive Neuroscience, 20*(12), 2125–2136. https://doi.org/10.1162/jocn.2008.20148

Isaac, R. M., & Walker, J. M. (1988). Group size effects in public goods provision: The voluntary contributions mechanism. *The Quarterly Journal of Economics, 103*(1), 179–199. https://doi.org/10.2307/1882648

Izuma, K., Saito, D. N., & Sadato, N. (2008). Processing of social and monetary rewards in the human striatum. *Neuron, 58*(2), 284–294. https://doi.org/10.1016/j.neuron.2008.03.020

Johnson, N. D., & Mislin, A. A. (2011). Trust games: A meta-analysis. *Journal of Economic Psychology, 32*(5), 865–889. https://doi.org/10.1016/j.joep.2011.05.007

Kahneman, D., Knetsch, J. L., & Thaler, R. (1986). Fairness as a constraint on profit seeking: Entitlements in the market. *The American Economic Review, 76*(4), 728–741.

King-Casas, B., Tomlin, D., Anen, C., Camerer, C. F., Quartz, S. R., & Montague, P. R. (2005). Getting to know you: Reputation and trust in a two-person economic exchange. *Science, 308*(5718), 78–83. https://doi.org/10.1126/science.1108062

Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V., & Fehr, E. (2006). Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science, 314*(5800), 829–832. https://doi.org/10.1126/science.1129156

Krueger, F., & Grafman, J. (2008). The human prefrontal cortex stores structured event complexes. In T. Shipley & J. M. Zacks (Eds.), *Understanding events: From perception to action* (pp. 617–638). Oxford University Press.

Krueger, F., Moll, J., Zahn, R., Heinecke, A., & Grafman, J. (2007). Event frequency modulates the processing of daily life activities in human medial prefrontal cortex. *Cerebral Cortex, 17*(10), 2346–2353. https://doi.org/10.1093/cercor/bhl143

Krupka, E. L., & Weber, R. A. (2013). Identifying social norms using coordination games: Why does dictator game sharing vary? *Journal of the European Economic Association, 11*(3), 495–524. https://doi.org/10.1111/jeea.12006

Larney, A., Rotella, A., & Barclay, P. (2019). Stake size effects in ultimatum game and dictator game offers: A meta-analysis. *Organizational Behavior and Human Decision Processes, 151*, 61–72. https://doi.org/10.1016/j.obhdp.2019.01.002

Ledyard, J. O. (1995). *Public goods: A survey of experimental research*. https://authors.library.caltech.edu/records/6w2tn-cpt82.

Lee, D. (2013). Decision making: From neuroscience to psychiatry. *Neuron, 78*(2), 233–248. https://doi.org/10.1016/j.neuron.2013.04.008

Lockwood, P. L., & Wittmann, M. K. (2018). Ventral anterior cingulate cortex and social decision-making. *Neuroscience & Biobehavioral Reviews, 92*, 187–191. https://doi.org/10.1016/j.neubiorev.2018.05.030

Lockwood, P. L., Apps, M. A. J., Valton, V., & Roiser, J. P. (2016). Neurocomputational mechanisms of prosocial learning and links to empathy. *Proceedings of the National Academy of Sciences, 113*(35), 9763–9768. https://doi.org/10.1073/pnas.1603198113

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

Mende-Siedlecki, P., Baron, S. G., & Todorov, A. (2013). Diagnostic value underlies asymmetric updating of impressions in the morality and ability domains. *Journal of Neuroscience, 33*(50), 19406–19415. https://doi.org/10.1523/jneurosci.2334-13.2013

Mengel, F. (2018). Risk and temptation: A meta-study on prisoner's dilemma games. *The Economic Journal, 128*(616), 3182–3209. https://doi.org/10.1111/ecoj.12548

Moll, J., Krueger, F., Zahn, R., Pardini, M., de Oliveira-Souza, R., & Grafman, J. (2006). Human fronto–mesolimbic networks guide decisions about charitable donation. *Proceedings of the National Academy of Sciences, 103*(42), 15623–15628. https://doi.org/10.1073/pnas.0604475103

Montague, P. R., Dolan, R. J., Friston, K. J., & Dayan, P. (2012). Computational psychiatry. *Trends in Cognitive Sciences, 16*(1), 72–80. https://doi.org/10.1016/j.tics.2011.11.018

Murnighan, J. K., Oesch, J. M., & Pillutla, M. (2001). Player types and self-impression Management in Dictatorship Games: Two experiments. *Games and Economic Behavior, 37*(2), 388–414. https://doi.org/10.1006/game.2001.0847

Nowak, M. A., Page, K. M., & Sigmund, K. (2000). Fairness versus reason in the ultimatum game. *Science, 289*(5485), 1773–1775.

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences, 105*(32), 11087–11092. https://doi.org/10.1073/pnas.0805664105

Peysakhovich, A., Nowak, M. A., & Rand, D. G. (2014). Humans display a 'cooperative phenotype' that is domain general and temporally stable. *Nature Communications, 5*(1), 4939. https://doi.org/10.1038/ncomms5939

Phan, K. L., Sripada, C. S., Angstadt, M., & McCabe, K. (2010). Reputation for reciprocity engages the brain reward center. *Proceedings of the National Academy of Sciences, 107*(29), 13099–13104. https://doi.org/10.1073/pnas.1008137107

Pletzer, J. L., Balliet, D., Joireman, J., Kuhlman, D. M., Voelpel, S. C., & Van Lange, P. A. M. (2018). Social Value Orientation, Expectations, and Cooperation in Social Dilemmas: A Meta–Analysis. European Journal of Personality, 32(1), 62–83. https://doi.org/10.1002/per.2139.

Pruitt, D. G. (1967). Reward structure and cooperation: The decomposed prisoner's dilemma game. *Journal of Personality and Social Psychology, 7*(1, Pt.1), 21–27. https://doi.org/10.1037/h0024914

Rhoads, S. A., Cutler, J., & Marsh, A. A. (2021). A feature-based network analysis and fMRI meta-analysis reveal three distinct types of prosocial decisions. *Social Cognitive and Affective Neuroscience, 16*(12), 1214–1233. https://doi.org/10.1093/scan/nsab079

Rilling, J. K., & Sanfey, A. G. (2011). The neuroscience of social decision-making. *Annual Review of Psychology, 62*, 23–48. https://doi.org/10.1146/annurev.psych.121208.131647

Rilling, J. K., Gutman, D. A., Zeh, T. R., Pagnoni, G., Berns, G. S., & Kilts, C. D. (2002). A neural basis for social cooperation. *Neuron, 35*(2), 395–405. https://doi.org/10.1016/S0896-6273(02)00755-9

Rilling, J. K., Goldsmith, D. R., Glenn, A. L., Jairam, M. R., Elfenbein, H. A., Dagenais, J. E., Murdock, C. D., & Pagnoni, G. (2008). The neural correlates of the affective response to unreciprocated cooperation. *Neuropsychologia, 46*(5), 1256–1266. https://doi.org/10.1016/j.neuropsychologia.2007.11.033

Ruff, C. C., Ugazio, G., & Fehr, E. (2013). Changing social norm compliance with noninvasive brain stimulation. *Science, 342*(6157), 482–484. https://doi.org/10.1126/science.1241399

Sally, D. (1995). Conversation and cooperation in social dilemmas: A meta-analysis of experiments from 1958 to 1992. *Rationality and Society, 7*(1), 58–92. https://doi.org/10.1177/1043463195007001004

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision-making in the ultimatum game. *Science, 300*(5626), 1755–1758. https://doi.org/10.1126/science.1082976

Sanfey, A. G., Loewenstein, G., McClure, S. M., & Cohen, J. D. (2006). Neuroeconomics: Cross-currents in research on decision-making. *Trends in Cognitive Sciences, 10*(3), 108–116. https://doi.org/10.1016/j.tics.2006.01.009

Sazhin, D., Wyngaarden, J. B., Dennison, J. B., Zaff, O., Fareri, D., McCloskey, M. S., Alloy, L. B., Jarcho, J. M., & Smith, D. V. (2024). Trait reward sensitivity modulates connectivity with the temporoparietal junction and anterior insula during strategic decision making. *bioRxiv*. https://doi.org/10.1101/2023.10.19.563125

Schreuders, E., Klapwijk, E. T., Will, G. J., & Crone, E. A. (2018). Friend versus foe: Neural correlates of prosocial decisions for liked and disliked peers. *Cognitive, Affective, & Behavioral Neuroscience, 18*(1), 127–142. https://doi.org/10.3758/s13415-017-0557-1

Schultz, W. (2002). Getting formal with dopamine and reward. *Neuron, 36*(2), 241–263. https://doi.org/10.1016/s0896-6273(02)00967-4

Schultz, W. (2013). Updating dopamine reward signals. *Current Opinion in Neurobiology, 23*(2), 229–238. https://doi.org/10.1016/j.conb.2012.11.012

Schulz, J. F., Fischbacher, U., Thöni, C., & Utikal, V. (2014). Affect and fairness: Dictator games under cognitive load. *Journal of Economic Psychology, 41*, 77–87. https://doi.org/10.1016/j.joep.2012.08.007

Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R. J., & Frith, C. D. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science, 303*(5661), 1157–1162. https://doi.org/10.1126/science.1093535

Singer, T., Seymour, B., O'Doherty, J. P., Stephan, K. E., Dolan, R. J., & Frith, C. D. (2006). Empathic neural responses are modulated by the perceived fairness of others. *Nature, 439*(7075), 466–469. https://doi.org/10.1038/nature04271

Spitzer, M., Fischbacher, U., Herrnberger, B., Grön, G., & Fehr, E. (2007). The neural signature of social norm compliance. *Neuron, 56*(1), 185–196. https://doi.org/10.1016/j.neuron.2007.09.011

Stallen, M., & Sanfey, A. G. (2013). The cooperative brain. *The Neuroscientist, 19*(3), 292–303. https://doi.org/10.1177/1073858412469728

Stallen, M., & Sanfey, A. G. (2015). The neuroscience of social conformity: Implications for fundamental and applied research. *Frontiers in Neuroscience, 9*, 337. https://doi.org/10.3389/fnins.2015.00337

Suzuki, S., Niki, K., Fujisaki, S., & Akiyama, E. (2011). Neural basis of conditional cooperation. *Social Cognitive and Affective Neuroscience, 6*(3), 338–347. https://doi.org/10.1093/scan/nsq042

Thompson, K., Nahmias, E., Fani, N., Kvaran, T., Turner, J., & Tone, E. (2021). The prisoner's dilemma paradigm provides a neurobiological framework for the social decision cascade. *PLoS One, 16*(3), e0248006. https://doi.org/10.1371/journal.pone.0248006

Todorov, A., & Engell, A. D. (2008). The role of the amygdala in implicit evaluation of emotionally neutral faces. *Social Cognitive and Affective Neuroscience, 3*(4), 303–312. https://doi.org/10.1093/scan/nsn033

Todorov, A., Pakrashi, M., & Oosterhof, N. N. (2009). Evaluating faces on trustworthiness after minimal time exposure. *Social Cognition, 27*(6), 813–833. https://doi.org/10.1521/soco.2009.27.6.813

Tomasello, M., & Vaish, A. (2013). Origins of human cooperation and morality. *Annual Review of Psychology, 64*, 231–255. https://doi.org/10.1146/annurev-psych-113011-143812

Tricomi, E., Rangel, A., Camerer, C. F., & O'Doherty, J. P. (2010). Neural evidence for inequality-averse social preferences. *Nature, 463*(7284), 1089–1091. https://doi.org/10.1038/nature08785

van Baar, J. M., Chang, L. J., & Sanfey, A. G. (2019). The computational and neural substrates of moral strategies in social decision-making. *Nature Communications, 10*(1), 1483. https://doi.org/10.1038/s41467-019-09161-6

van't Wout, M., Kahn, R. S., Sanfey, A. G., & Aleman, A. (2005). Repetitive transcranial magnetic stimulation over the right dorsolateral prefrontal cortex affects strategic decision-making. *Neuroreport, 16*(16), 1849–1852. https://doi.org/10.1097/01.wnr.0000183907.08149.14

Vives, M.-L., & FeldmanHall, O. (2018). Tolerance to ambiguous uncertainty predicts prosocial behavior. *Nature Communications, 9*(1), 2156. https://doi.org/10.1038/s41467-018-04631-9

von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior* (pp. xviii–625). Princeton University Press.

Wells, J., & Rand, D. G. (2013). *Strategic self-interest can explain seemingly "fair" offers in the ultimatum game* (SSRN scholarly paper no. 2136707). https://doi.org/10.2139/ssrn.2136707

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science, 17*(7), 592–598. https://doi.org/10.1111/j.1467-9280.2006.01750.x