Computations in the social brain
Rojek-Giffin, M.

Author: Rojek-Giffin, M.
Title: Computations in the social brain
Issue Date: 2021-05-26

Computations in the Social Brain

Michael Rojek-Giffin

ISBN: xxxxxxxxxxxxxx
Cover artwork: Shiro36, based on art by Josan Gonzalez
Lay-out: Dennis Hendriks / ProefschriftMaken.nl
Printing: ProefschriftMaken.nl

Computations in the Social Brain

Proefschrift

ter verkrijging van
de graad van doctor aan de Universiteit Leiden,
op gezag van rector magnificus prof.dr.ir. H. Bijl,
volgens besluit van het college voor promoties

te verdedigen op woensdag 26 mei 2021
klokke 15.00 uur

by

Michael Rojek-Giffin
geboren te de Verenigde Staten
in 1988

Promotor
Prof. dr. C. K. W. de Dreu (Leiden University)

Co-Promotor
Dr. Jörg Gross (Leiden University)

Doctorate Committee
Prof. dr. Eric van Dijk (Leiden University)
Dhr. dr. J.B. (Jan) Engelmann (University of Amsterdam)
Prof. dr. Marie Claire Villeval (University of Lyon)
Prof. dr. A.G. Sanfey (Radboud University)
Dr. Mirre Stallen (Leiden University)
Prof. dr. S.T. Nieuwenhuis (Leiden University)

Table of Contents

1

Chapter 1

Introduction

We respond to gestures with an extreme alertness and, one might almost say, in accordance with an elaborate and secret code that is written nowhere, known by none, and understood by all.

Edward Sapir, 1927

The social animal

Humans rely on one another for protection against outside danger (De Dreu & Gross, 2019; Tomasello, 2019), for personal meaning (Cacioppo & Hawkley, 2009; Ortner, 1998), and for advancing technological innovation and cultural evolution (De Dreu, Nijstad, Bechtoldt, & Baas, 2011; Henrich, 2016). However, despite the benefits afforded by living in social groups and cooperating with others, these benefits can come at a cost to the individuals within the group. Oftentimes, the actions most beneficial to the group and the actions most beneficial to the individuals within that group are incongruent, and cooperative behavior requires an individual to forego a personal benefit for the sake of their social unit. It would be more personally profitable to neglect the bill after dining out, to free-ride on public transport, and to claim ownership over a valuable idea without giving proper credit. However, if everyone adhered to this self-interested ethos society could not function, and no one would be allowed to enjoy the benefits of the public goods that civilization provides.

Indeed, society could not function if it were not for the fact that, as Jean-Jacques Rousseau put it (1762/1993): “each of us puts his person ... under the supreme direction of the general will” (p. 196). While this statement is true in an ideal world, the danger that a customer does not pay their bill after dining out, that a passenger freely partakes in public transport, or that a colleague claims another’s ideas as their own, is always a possibility. Cooperating within the bounds of a group’s customs goes hand in hand with the risk that other group members act with purely self-interested motives, exploiting the behaviors of those under the “direction of the general will” (Rousseau, 1762/1993). In other words, in addition to the personal costs, there are *risks* of interpersonal exploitation inherent to social exchange. These risks are mitigated by specific institutions, such as norms of cooperation, which curb antisocial behavior (Bicchieri, 2005; Fehr & Schurtenberger, 2018). However, norms are an implicit set of rules – what the great anthropologist Edward Sapir (1927) referred to as “an elaborate and secret code” (p. 137), which means that norms can be ambiguous (e.g. knowing how much to tip in a restaurant), and therefore learning and applying norms is an elusive task. Adequately navigating a social group’s norms is both essential and challenging, and failure to do so is both easily accomplished and detrimental. Yet, humans somehow do learn to navigate these implicit rules that govern civil society.

Recent advances in psychology, economics, and neuroscience attempt to shed light on these problems by combining (i) structured situations with economic games in tightly controlled experimental setups, (ii) formal theory with the application of mathematical frameworks such as game theory and expected utility theory, and (iii) neuroscientific methods that allow us to assess hypothesized neurophysiological mechanisms. These approaches have in common that they assume that humans attempt to maximize some form of subjective value, so-called utility. In the remainder of this introduction, I will outline how these advances were put to use in the current thesis in order to elucidate some of the outstanding questions surrounding human social interactions.

The utility of economic games

One of the tools used to study how individuals navigate complex social interactions is Game Theory. Originally invented by mathematicians and physicists to study strategic interdependence (Dimand & Dimand, 1996), Game Theory offers a precise mathematical formulation of decision-making under clearly defined circumstances. A "game" in this framework is a specification of the strategies, information, outcomes, and associated values available to interacting players (Camerer, 2003). A canonical example is the prisoner's dilemma game, which in its simplest form involves two players, each of whom must decide whether to "cooperate" or "defect" with the other. If both players cooperate or both defect, they both receive equal payments, with mutual cooperation leading to higher payoffs than mutual defection. However, unilateral defection (in which one player exploits the other's cooperative decision) leads to asymmetrical payoffs benefitting the defector. This game offers a simple model allowing for the study of social uncertainty and strategic reasoning, such as deciding whether or not to pay the bill after dining out, pay for public transport, or claim ownership over a valuable idea without giving proper credit.

When considering that in virtually every social interaction, human individuals (*viz.* players) have available actions (*viz.* strategies) with associated outcomes (*viz.* payoffs), it becomes clear that quite a lot of human social life can be classified as a "game". Indeed, questions regarding who will attack a vulnerable opponent when competing for resources, or trust and reciprocate with an unknown other, or learn what's considered fair in a novel environment, all seem intractable at face value. However, each of these situations can be modeled by relatively simple economic games. Here I focus on three games in particular – the attacker-defender contest, the trust game, and the ultimatum game – which model asymmetrical conflicts, generosity and reciprocity, and norms of fairness, respectively.

The attacker-defender contest.

Failure to navigate a group's code of conduct can lead to interpersonal conflict. Conflict itself is a complex and multifaceted phenomenon, making a precise definition problematic. Nevertheless, conflict generally involves an incongruence of desires between multiple interaction parties. Several individuals striving for the last piece of cake, only parking space, or coveted job opening, all can result in conflict. In these examples (as in many cases), conflict can emerge out of a symmetrical structure, in which all players have the same goals, knowledge and available actions. However, most conflicts exhibit an asymmetry of power and motivation between the involved parties (De Dreu & Gross, 2019; De Dreu, Gross, et al., 2016). Indeed, roughly 67% of interstate militarized conflicts involve one revisionist state (one nation seeking change in another) and one non-revisionist state (one nation resisting change from another) (De Dreu, Gross, et al., 2016). This same asymmetry exists in corporate hostile takeovers (Schwert, 2000), in dissolution of romantic entanglements (Kluwer, Heesink, & Van de Vliert, 1997; Perel, 2017), and in groups of predators hunting prey (Dawkins & Krebs, 1979; De Dreu, Gross, et al., 2016).

To model the processes underpinning these asymmetric conflicts, we designed an economic game called the attacker-defender contest. This game consists of two players, each of whom starts with a given endowment out of which they can invest. One player can invest in attack, and the other player can invest in defense. Investments are non-recoverable and thus wasted. However, if the attacker invests more than the defender, then the attacker obtains all the leftover endowment of the defender, i.e. whatever the defender did not invest. If this happens, the defender ends with nothing. If, however, the defender invests as much or more than the attacker, both sides keep their non-invested resources. In this setup, investments can increase attacker earnings, and can prevent defenders from losing their remaining endowment to their attacker. Across multiple studies this game has revealed that investments in defense are more frequent and more intense than investments in attack (De Dreu, Giacomantonio, Giffin, & Vecchiato, 2019; De Dreu & Gross, 2019; De Dreu, Gross, et al., 2016). Furthermore, attacks are only successful (i.e. result in taking the defenders' remaining endowment) in about 30% of cases, which mimics the success rates in interstate warfare, corporate hostile takeovers, and group-hunting predators (De Dreu, Gross, et al., 2016).

The situation in which both attackers and defenders invest nothing is the situation that leads to the most collective wealth – they both keep their entire endowments. However, the attacker has the opportunity to earn more than their endowment if they invest. It is most beneficial for the attacker-defender *unit* to invest nothing, yet it is more individually beneficial for the attacker to invest and take the defender's endowment, allowing for a precise model of the mixed-motives structures alluded to above.

The trust game.

While conflict can result from an asymmetry of motives, it can also be sparked by a mismatch between what is promised or expected and what is delivered or experienced. This can lead to innocuous situations such as roommates squabbling over dirty dishes, and to severe situations such as interstate war. For example, during the revolutionary war of the United States, George Washington, who was the head of the American military, made an alliance with the King of France and accepted French aid in the fight against the British Empire. However, when the French people sought American military assistance during their own revolution, Washington claimed the alliance was with the King, not the people, and refused to provide any help. A similar example can be found in South America, where the ruthless conquistador Francisco Pizarro captured the Incan emperor Atahualpa and guaranteed his safety in exchange for gold. When Pizarro received his payment, he reneged on his promise, killing Atahualpa and continuing his subjugation of the Incan people. In these examples, understandings of what constituted a mutually trusting relationship between the parties involved were either at odds, or else one party violated said trust. The results in both cases was a severe hampering or complete destruction of future social interactions between the involved parties.

How and when individuals decide whether to trust others, and whether to reciprocate said trust, can be addressed with a variation of the trust game (Berg, Dickhaut, & McCabe, 1995). The trust game consists of a sender who decides how much (if any) of a given endowment to transfer (*viz.* entrust) to a responder. The amount transferred to the responder is then increased by some multiplying factor (usually three), after which the responder decides how much (if any) to return (*viz.* reciprocate) back to the sender. Because each unit of the endowment the sender transfers to the responder is tripled, the most collectively profitable outcome involves the sender transferring their entire endowment to the responder. However, in this situation the responder has a strong incentive to exploit the sender's trust and keep the entire sum for themselves. Therefore, the situation that creates the most collective wealth is also the situation that can create the largest inequity and risk of exploitation. So while perhaps the United States aiding France in their struggle for revolution or Pizarro releasing the Incan emperor as promised could have created the most collectively advantageous scenario, failing to reciprocate may have created the most individually advantageous scenario (or at least the expectation thereof) for the United States and Pizarro.

While it is certainly tempting to infer the motivations behind the dramatic actions of George Washington or Francisco Pizarro, a psychological account in these cases remains speculative. This is precisely why controlled laboratory experiments are so valuable. Previous research on behavior in the trust game shows that, on average, senders transfer half of their endowments to responders who, on average, return 40% of the tripled amount back to the sender (Johnson & Mislin, 2011), which implies that people are, for the most part, trusting and trustworthy. However, there is substantial inter-individual and cross-cultural variation in both trust and trustworthiness (Balliet & Van Lange, 2013; Bohnet, Greig, Herrmann, & Zeckhauser, 2008; Johnson & Mislin, 2011; Romano, Balliet, Yamagishi, & Liu, 2017), suggesting that different individuals follow different rules when deciding to trust and reciprocate. How individuals learn these rules is an open question in need of more research.

The ultimatum game.

One element contributing to differences of expectations are differences between cultures. Navigating the rules and secret codes of one's own culture is already a difficult task, and straddling multiple cultures only increases the difficulty. There is a multiplexity of customs, rituals, and taboos which threaten the efficacy of the interaction. Nevertheless, in the increasingly globalized world, navigating the cultural divide with little to no knowledge of other individuals' cultural backgrounds has become commonplace.

To model the mechanisms behind learning the implicit rules associated with different cultural expectations, we utilize the ultimatum game (Güth, Schmittberger, & Schwarze, 1982), a two-player game in which one player (the proposer) starts with an endowment from which they make an offer to the other player (the responder). The responder then decides whether to accept the proposed division, or to reject it, in which case both players receive nothing. A proposer ideally makes an offer that is exactly at the acceptability threshold of the responder. Offering too little risks rejection, but offering too much is an unnecessary expenditure. This game mimics the type of judgements involved in both innocuous and high-stakes interactions. For example, when a store owner is selecting an item's price, setting too high of a price risks scaring off potential customers, and setting too low of a price risks missing out on potential profits. A much more high-stakes example is the negotiation of a peace treaty – if one party demands terms that are too stringent, they risk enraging the other party and dissolving the agreement. However if one party demands terms that are too weak, they could be making unnecessary concessions.

Cross-cultural research with the ultimatum game have shown a diverse set of behaviors and norms across different parts of the world (Henrich et al., 2005; Henrich, Ensminger, et al., 2010). In our version, we experimentally nudge individuals to accept different offers, modeling different fairness norms, which we use to explore how individuals learn different implicit rules of engagement.

The correspondence problem.

Economic games provide stripped back models of complex behavior, and by truncating the number of options and outcomes available to players they limit the amount of motivations that could be driving the observed strategies. Nevertheless, there is not necessarily a single psychological trait responsible for every single action. In other words, there is not a one-to-one correspondence between an action and a purported psychological driver of said action. In the attacker-defender contest, for example, one player might invest in defense out of mistrust of the other player, to reduce their personal uncertainty about the game's outcome, or because they feel that this action is expected of them under the circumstances. Therefore, while economic games reduce the complexity of human social behavior to a more manageable set of variables, we are still inevitably left with ambiguity regarding the foundations of human behavior and concomitant cognitive processes. For this reason, precise quantification of behavior into a set of formal suppositions via computational modeling enhances our predictive ability regarding what drives individuals to act the way that they do.

Computational modeling

Computational models provide us with a mathematical language with which to make predictions about the mechanisms driving behavior. To employ a computational model, one formulates a series of calculations that are hypothesized to generate a given outcome, in our case a certain social propensity. While the term "model" can be used to describe any mathematical or even conceptual mapping of one phenomenon onto another, in our sense we mean specifically a series of equations which describe the mechanisms behind an observed behavior. In this way, the models which we employ are *generative models*: they are mathematical expressions of the processes which generate behavior.

While there are scores of different computational models that have been used to describe virtually any simple or complex system, from the weather (Lynch, 2008) to a chess match (Larson, 2010), from presidential elections (Silver, 2012) to the spread of the Coronavirus (Friston et al., 2020), what all these models have in common is that each of them consists of a set of parameters constituting algorithms which transform inputs (rules of the game, personality traits, economic status, dispositions) into outputs (actions, cooperating vs. defecting, attacking vs. abstaining). These parameters are simply numerical weights which are estimated from the data. More importantly, these parameters oftentimes come along with psychological interpretations. For example, the most widely used technique for modeling choice selection is the softmax choice function (Daw, 2011; Sutton & Barto, 2018). This function consists of a single free (i.e. estimated) parameter, the so-called inverse temperature parameter. A high value for this parameter denotes high determinism, indicating that the individual is exploiting a known course of action instead of exploring unknown courses. A low value for this parameter denotes high stochasticity, indicating that the individual is exploring unknown options. Put succinctly, this parameter gives us a precise measure of an individual's exploration/exploitation tradeoff. It is precisely this type of translation from numerical precision to qualitative psychological interpretation that make computational models such a powerful tool for studying human behavior.

As stated above, the number of possible computational models that exist is staggering (Jolly & Chang, 2019; Palminteri, Wyart, & Koechlin, 2017; Sutton & Barto, 2018). However, much research on behavior in economic games has found consistent and robust evidence in favor of a computational framework that is based on the assumption that the individual is attempting to maximize subjective value, or utility

(Camerer, 2003). The concept of utility has many definitions (Georgescu-Roegen, 1968), but perhaps the famous philosopher Jeremy Bentham, the founder of utilitarianism, put it best: "nature has placed mankind under the governance of two sovereign masters, pain and pleasure… They govern us in all we do, in all we say, in all we think… The principle of utility recognizes this subjection, and assumes it for the foundation of that system" (1789). Put succinctly, each of our models assume that individuals are attempting to maximize pleasure and minimize pain, which we operationalize as maximizing collective or individual reward, and minimizing collective or individual loss.

Based on this principle, we still have a potential multitude of models at our disposal, however each makes use of the concept of utility and, as such, each facilitates the use of *prediction errors*. A prediction error is simply a measure of the mismatch between what is expected and what is actually experienced. Learning itself is essentially reducing prediction errors over time (Sutton & Barto, 2018), and this principle is a powerful tool with which to study adaptive behavior. Furthermore, predictions errors have a robust and consistent fingerprint in the brain, which allows for a way to connect cognitive and neurological processes.

Neuroimaging

Combining neuroimaging with economic games and computational modeling increases our ability to explain the psychological and concomitant neurophysiological substrates underlying behavior. The neuroimaging method most commonly used in conjunction with economic games and computational modeling, and the one we employ here, is functional magnetic resonance imaging (fMRI) (Behrens, Hunt, & Rushworth, 2009; Fehr & Camerer, 2007).

In fMRI, a participant completes a task (such as an economic game) while lying supine in an MRI scanner. Throughout the task, the MRI scanner uses magnetic waves to take precise measures of the blood flow in the participant's brain. Importantly, blood has a specific magnetic signature depending on how much oxygen it contains. Because more cellular activity requires more oxygen, we can use the magnetic signatures of oxygenated vs. deoxygenated blood to infer levels of neural activity. In fact, research has shown that there is a close correspondence between neural activity and changes in blood oxygenation levels. This allows the use of blood oxygen level dependent changes, so called BOLD responses, as measured by fMRI, as proxies of the underlying neural activity (Logothetis, Pauls, Augath, Trinath, & Oeltermann, 2001).

There has, however, been controversy in recent years surrounding the efficacy of fMRI. Specifically, some studies have shown that, using conventional analysis techniques, a statistically significant signal can be detected when in reality no signal is present (Bennett, Miller, & Wolford, 2009; Eklund, Nichols, & Knutsson, 2016; Warren et al., 2017). Furthermore, the very practice of fMRI often makes use of "reverse inference", in which a psychological construct is inferred based on the activity of a certain brain region. This type of logic, though widely practiced, is oftentimes invalid (Poldrack, 2006, 2011). However, this type of fallacious logic can be avoided by selecting neural structures a priori and using independent tasks that elicit activations in these same areas in order to localize the subject-specific functional regions (e.g., Prochazkova et al., 2018).

Furthermore, one of the most reliable and robust effects within cognitive neuroscience is the discovery of so-called reward prediction errors in the brain (Gershman & Uchida, 2019; Preuschoff & Bossaerts, 2007; Schultz, 2010; Schultz, Dayan, & Montague, 1997). In a seminal study (Schultz et al., 1997), monkeys were trained to expect a reward (juice) every time they pulled a lever. At the beginning of the learning process, each time a monkey pulled the lever and received juice, neurons within the dopaminergic midbrain became more active. This is consistent with the interpretation that the monkeys were surprised each time a lever pull resulted in a reward; they experienced a positive prediction error, i.e. the outcome was better than they expected, and this positive prediction error was expressed as an increase in dopaminergic neural activity. When, on the other hand, the monkeys had experienced that a lever pull preceded juice allocation many times, yet no juice was provided, neurons within the dopaminergic midbrain became less active. This is consistent with the interpretation that the monkeys were still surprised, but this time from a negative prediction error – the outcome was worse than they expected – and this negative prediction error was expressed as a decrease in dopaminergic neural activity. This close coupling of prediction errors with neural firing patterns was subsequently shown to be robustly correlated with BOLD response within regions of the human basal ganglia, most notably the ventral striatum and ventromedial prefrontal cortex (Behrens, Hunt, Woolrich, & Rushworth, 2008; Behrens, Woolrich, Walton, & Rushworth, 2007; O'Doherty et al., 2004), which is heavily innervated with dopaminergic inputs (Palminteri & Pessiglione, 2017).

Importantly, this close correspondence between neural activity and subjective experience (e.g. surprise) facilitates the use of computational models that make use of the concept of utility. Specifically, when a participant plays an economic game in the fMRI scanner, we use computational modeling to infer their expectations and violations thereof; in other words, we estimate their predictions and their prediction errors. We can then search for correlates of these prediction errors in the brain. In fact, because the neural reward prediction error is so reliable, we are able to use neural data to validate our computational models. In short, economic games, computational modeling, and fMRI each strengthen each other in a three-pronged approach to the study human sociality.

Outline of thesis

This thesis consists of three empirical chapters that investigate elements of human social behavior through the combination of economic games, computational modeling, and neuroimaging. The empirical chapters are outlined in detail below.

Chapter 2.

Chapter 2 deals with the assertion made by John Stuart Mill (1859) in his *principles of political economy*: "a great proportion of all efforts … [are] spent by mankind in injuring one another, or in protecting against injury." These tendencies for "injuring others" and defending against injury are well captured by economic contest experiments, such as our attacker-defender contest, in which individuals invest to gain a reward at a cost to their competitor (so-called attack), or to avoid losing their resources to their antagonist (Carter & Anderton, 2001; Chen & Bao, 2015; Chowdhury, Jeon, & Ramalingam, 2018; De Dreu & Gross, 2019; De Dreu, Kret, & Sligte, 2016; De Dreu, Scholte, van Winden, & Ridderinkhof, 2015; Grossman & Kim, 1996; Wittmann et al., 2016; Zhu, Mathewson, & Hsu, 2012).

Why individuals invest in attack and defense remains poorly understood, and can be explained by a variety of subjective "desires" (Charpentier, Aylward, Roiser, & Robinson, 2017; Delgado, Schotter, Ozbay, & Phelps, 2008; Dorris & Glimcher, 2004). Humans may be attempting to maximize their personal earnings when investing in attack and defense, a desire which is typically assumed in standard economic theory (e.g. Ostrom, 1998). In the same vein, individuals could invest in attack and defense due to "competitive arousal" and interpersonal rivalry (Delgado et al., 2008; Ku, Malhotra, & Murnighan, 2005). Furthermore, attack and defense investments could be indicative of a desire to minimize risk and uncertainty (Delgado et al., 2008; Kahneman & Tversky, 1984). In short, behavior in this game can be driven by a multitude of psychological forces.

We addressed this multiplicity of motives using an approach in line with research on learning from reward and risk prediction (Olsson, FeldmanHall, Haaker, & Hensler, 2018; Palminteri, Wyart, & Koechlin, 2017; Preuschoff, Quartz, & Bossaerts, 2008). In conjunction with fMRI, we applied a cognitive-hierarchies framework (Camerer, Ho, & Chong, 2004). The cognitive-hierarchies framework rests on the assumption that expectations and beliefs in strategic interactions are formed recursively (i.e., [1] I think that [2] you think that [3] I think that [4]…) and vary in terms of their sophistication (i.e., the number of recursions *k*). Using this computational framework, we were able to estimate, for each attack/defense investment, the expected reward and accompanying prediction errors.

Our results showed that attackers were best described by a model with 4 levels of recursion, while defenders were best described by a model with 3 levels of recursion. This suggests that during attack individuals engage a more sophisticated level of recursive reasoning than during defense. At the neural level we found that participants exhibited more neural activity during attack relative to defense in the anterior insula, a region associated with emotional processing (Sanfey, Rilling, Aronson, Nystrom, & Cohen, 2003), and the inferior frontal gyrus, a region associated with strategic reasoning (De Dreu, Kret, et al., 2016) as well as theory of mind (Engelmann, Meyer, Ruff & Fehr, 2019; Prochazkova et al., 2018; Van Overwalle, 2009). In a follow-up analysis we found that neural activity during attack covaried with wins and losses in, among other regions, the temporoparietal junction, a region consistently linked to theory of mind (Engelmann, Meyer, Ruff & Fehr, 2019; Prochazkova et al., 2018; Van Overwalle, 2009), and the ventral striatum, one of the central hubs of the reward learning network (Balodis et al., 2012; McNamee, Rangel, & O'Doherty, 2013; Metereau & Dreher, 2015; Rudorf, Preuschoff, & Weber, 2012; Xue et al., 2009; Zhu et al., 2012). Within the ventral striatum, there was a significantly higher correlation between reward prediction errors and neural activity during attack than during defense, suggesting that attacker brains were more responsive to the rewards of the contest than defender brains.

In sum, our results suggest that strategic injuring of others is accomplished via high-level recursive reasoning with the goal of maximizing personal wealth. This is given credence by the fact that attackers utilized higher *k*-level reasoning than did defenders, as well as the fact that neural structures associated with theory of mind and reward processing were preferentially recruited during attack relative to defense. Taken together, our results suggests that theory of mind, while essential for empathy, could also underpin strategies which serve to maximize reward through exploiting and subordinating others.

Chapter 3.

A key task for defenders in the attacker-defender game studied in Chapter 2 is to assess to what extent they can trust their counterpart to not attack, or should instead fear their counterpart's aggressiveness. Chapter 3 zooms in on trust and distrust as a key element in social interactions. Especially when interacting with strangers, the decision to trust is non-trivial, as norms of trust and reciprocity differ dramatically between cultures and groups (Heap & Zizzo, 2009; Johnson & Mislin, 2011; Romano, Balliet, & Wu, 2017; Romano, Balliet, Yamagishi, et al., 2017). Therefore, learning these norms in novel situations is critically important for individuals to adequately function within a social environment. Faulty predictions of an interlocutor's norms can lead to losing out, either from refusal to engage in a mutually trusting relationship, or from engaging in a relationship that results in exploitation. Accurate predictions, on the other hand, allow an individual to distinguish the trustworthy from the exploitative. It follows that, individuals require the ability to learn to predict the reciprocity of others. We examine this supposition by applying computational modeling to behavior in a variation of the trust game.

Previous research on the trust game suggests that different individuals operate under different sets of rules regarding trust. Indeed, in this chapter we uncover that nearly all individuals fall into one of three discrete categories of reciprocity: exploiters, perfect reciprocators, and contingent reciprocators. Exploiters are responders who never return as much money to the sender as the sender transferred to them. Perfect reciprocators are responders who always return at least as much money as the sender transferred to them. Contingent reciprocators are responders who return money as a function of how much the sender transferred to them – when the sender transfers a small amount, they return a small amount, and when the sender transfers a large amount, they return a large amount.

These different reciprocity types raise an important question: how do individuals learn who to trust and who to avoid? In order to address this, we confronted naïve individuals with these different types, and gave them the opportunity to learn the trustworthiness of each through repeated interactions. We then constructed several computational learning models. One way in which individuals could learn who to trust and who to avoid is through the use of reinforcement learning (RL). At its simplest, RL makes the claim that individuals make predictions about actions and then update those predictions based on the observed outcomes (Sutton & Barto, 2018). This framework has been effective in capturing a wide variety of behaviors (Behrens et al., 2009; Erev & Roth, 1998; Palminteri et al., 2017), and precisely predicts the neural correlates of the learning process (Behrens et al., 2008; Levy & Glimcher, 2012; Rutledge, Dean, Caplin, & Glimcher, 2010). However, RL does make some psychologically implausible assumptions about how individuals reason. For example, RL only updates values for the selected action. This means that, under RL, we assume individuals ignore the value of every action except for the one they choose.

In a different computational account, individuals mentally simulate the outcome of all actions, and update the values of these actions accordingly. This so-called belief-based learning (BB) expands RL to include beliefs and counterfactual action simulation. This framework, however, also suffers from some untenable assumptions. For example, all outcomes, for both selected and mentally simulated actions, are updated with equal weight. This means that imagined and experienced outcomes are treated identically.

Both of these frameworks have been combined in a hybrid model, Experience Weighted Attraction (EWA), which, better than RL or BB accounts, describes behavior in a variety of economic games such as the trust game (Camerer & Ho, 1999; Camerer, Ho, & Chong, 2002; Zhu et al., 2012) . Moreover, certain parameters of the EWA model have exact psychological interpretations, which further facilitates inferences about the processes generating behavior.

When applied to the behavior of individuals playing the trust game as senders against the three aforementioned responder categories (exploiter, perfect reciprocator, contingent reciprocator), we found that the EWA model captured behavior better than

RL and BB alternatives. This means that when individuals are learning who they can trust and who they cannot, they combine their own experiences with their personal beliefs in a hybrid fashion. Interestingly, subjects behaved sub-optimally against all the different responder types, especially when interacting with contingent reciprocators. We further showed that the degree to which individuals learned from their simulated outcomes, the more money they earned from their interactions. This indicates that mental simulation while learning to trust offers a tangible benefit to the individual.

In sum, we show that people cannot necessarily be categorized as simply trustworthy or untrustworthy because a substantial proportion of individuals reciprocate trust based on how much they are trusted. Furthermore, we show that to learn these different trustworthy categories individuals employ a combination of experiential and belief-based learning. Taken together this suggests that to effectively learn who is trustworthy and who is not individuals must form a strong internal concept of the social world.

Chapter 4 .

The results from both Chapter 2 and 3 revealed an important role for social perception and learning, suggesting that empathy and social norms modulate decisions to exploit and to trust and reciprocate. Chapter 4 builds on these and related findings by asking what role empathy (Zaki, 2014; Zaki & Mitchell, 2013) and social preferences such as concerns for fairness and the welfare of others (Blake et al., 2015; Fehr & Schmidt, 1999) play in learning group-specific conventions. Do these traits facilitate social coordination by enabling the acquisition of culture-specific rules of engagement, or do they interfere with efficient learning by biasing beliefs and hindering accurate information updating? Existing theory is ill-suited to answer these questions. Specifically, the role of social preferences in the formation and updating of beliefs and expectations is poorly understood. For this reason, we investigated whether and how beliefs about others' needs and desires are formed and updated, and whether and how culturally engrained social preferences shape the learning of rules of engagement.

In a first step, we created three distinct groups of ultimatum game responders that differed in the extent to which individual members would accept (versus reject) ultimatum offers based on actual decisions from participants. These different groups exhibited different acceptance thresholds, i.e. the minimum offer they would accept. This amount was unknown to proposers, yet each different group was identified with a unique symbol, similar to culture-specific markers of identity such as language or clothing. In addition to the three different groups of responders, we also tested whether social consequences affected the degree to which proposers learned culture-specific rules of engagement. In one treatment (the social condition), proposers interacted with human responders whose earnings depended on the (acceptance of) proposed offers. In the other treatment (the non-social condition), proposers interacted with behaviorally identical computer agents that did not earn from the (acceptance of) proposed offers

(Baumgartner, Fischbacher, Feierabend, Lutz, & Fehr, 2009; Sanfey et al., 2003). In the social condition, participants were explicitly told that they were facing groups of responders who had received different starting endowments but not what the endowments were. In the non-social condition, participants were told that they were facing computer generated lotteries programmed to mimic the behavior of participants who had received different starting endowments. This created a social and non-social learning environment with identical learning contingencies, allowing us to test how social concerns for the responders affects learning.

We next constructed a computational framework to model how proposers behave when paired with responders from these different groups, as well as how proposers behave when interacting in a social versus a non-social context. We used a so-called Bayesian Preference Learner (BPL) model, which posits that individuals learn by applying Bayes' theorem. Through simulations we demonstrated that this model effectively captures optimal learning of these different responder groups in our ultimatum game setup. Furthermore, we demonstrated through simulations that learning should differ between social and non-social contexts if we assume that individuals exhibit an aversion to unequal outcomes for themselves versus other individuals – so-called inequity aversion.

At the behavioral level we found that proposers did indeed learn the different responder groups, and that this learning process was well captured by our BPL model. This was likewise corroborated by an fMRI analysis, wherein we found prediction errors from our BPL model in the ventral striatum and ventromedial prefrontal cortex, crucial hubs in the reward learning network (Balodis et al., 2012; McNamee et al., 2013; Metereau & Dreher, 2015; Rudorf et al., 2012; Xue et al., 2009; Zhu et al., 2012) We also found that proposers made higher offers to responders in the social relative to the non-social condition, and that this discrepancy between conditions could be explained by including an inequity aversion term in our BPL model. This social/non-social difference was also expressed in the brain, with significant differences in the dorsal anterior cingulate cortex, a structure consistently associated with executive control and contextual updating (Ebitz & Hayden, 2016; Kolling et al., 2016; Meder et al., 2016), as well as several hubs of the theory of mind network such as the superior temporal sulcus and the precuneus (Coricelli & Nagel, 2009; Hampton, Bossaerts, & O'Doherty, 2008). Furthermore, in a post-hoc belief estimation task, we found that proposers actually over-estimated the acceptance thresholds of responders in the social relative to the non-social condition. This suggests that social concerns and personal preferences modify not only how individuals behave but what they believe about others in their social environment, which in turn biases how they build and update their concepts of culture-specific rules of engagement.

In sum, this chapter shows that humans can learn the cultural norm of an environment through a process of Bayesian learning, relying on reward/reinforcement neural circuitry. This process, however, is hindered by moral sentiments and social preferences, which lead individuals to rely on erroneous heuristics, make unnecessarily high offers to their opponents, and leave money on the table. However, while this process leads to false beliefs about the environment and facilitates the perpetuation of relying on erroneous heuristics, the resulting self-fulfilling prophecy (I believe others require me to be nice and therefore I am nice) may explain how norms of fairness can establish and survive in groups even when individuals have selfish motives.

Conclusions

While each empirical chapter uses a different computational model applied to a different economic game, each chapter addresses a specific problem regarding human social living. In this way, each chapter contributes to a greater understanding of how humans are able to socially interact despite the incentives which repel individuals away from commonality. Through the attacker-defender contest we show that when motives and abilities are asymmetrical such that one party (the attacker) can benefit at the expense of the other (the defender), attackers will utilize high-level recursive reasoning and associated neural circuitry in an attempt to profit at the expense of defenders. Through the trust game we show that reciprocity behavior falls in three simple categories, yet learning these categories occurs suboptimally through a combination of belief and experiential learning – with greater reliance on personal belief being associated with higher monetary outcomes. Through the ultimatum game we show that individuals can be nudged into exhibiting acceptability thresholds that resemble those of different cultures. Learning to adapt to these different "cultures" is impeded by social preferences, which in turn leads to false beliefs about the social environment. Each study deals with a different collection of social norms and attempts to understand how humans reconcile the associated social dilemmas.

While each chapter does contribute to a more comprehensive understanding of human sociality, there are shortcomings to the included works that can be improved upon in future studies. One of the largest outstanding questions concerns generalizability. All of our participants came from western, educated, industrialized, rich, and democratic backgrounds – so-called WEIRD societies (Henrich, Heine, & Norenzayan, 2010). It could very well be the case that the effects found in our studies are specific to this particular milieu. Future studies involving cross-cultural comparisons are needed in order to establish how generalizable our results truly are. Furthermore, while we do attempt to describe in detail the mechanisms generating the behaviors we observe, there are many outstanding questions regarding environmental, psychological, and neurological substrates.

In the attacker-defender contest, future studies should attempt to promote greater cooperation between the interacting parties. Is there a particular psychological framing, such as stressing similarity between the attacker and defender, that will prevent attackers and defenders from investing in the contest in order to maximize collective wealth? Are there some changes to the incentive structure that can turn a docile attacker into an aggressive attacker (or vice versa)? In our study, attack behavior was associated with more sophisticated recursive reasoning. Could we potentially nudge individuals into higher (or lower) levels of recursion, and if so would that make them more (or less) efficacious in this setup? We also show that during attack relative to defense, reward related neural circuitry was preferentially responsive; this might suggest that dampening the effects of key neurotransmitters in the reward system (e.g. dopamine) might also reduce attack behavior.

In the trust game, to our knowledge no other research has discovered the existence of the sizeable minority of responders we call contingent reciprocators. An interesting follow-up question should address what traits predict whether an individual will fall into this category versus the other two. Are contingent reciprocators more calculating? More selfish? More future-oriented? Is position in one of these three categories stable over time, or do individuals shift their reciprocity behavior frequently? Another interesting question that arises is: despite the fact that these reciprocity types seem ubiquitous, why do individuals learn them so imperfectly? Can we introduce manipulations, such as observational learning, in order to facilitate optimal learning of these types? Finally, future studies should elucidate what neural correlates underpin this learning process. Is the suboptimality of learning associated with diminished activity in reward learning circuitry, or associated with regions more involved in theory of mind and social cognition? Answering these questions could help determine what aspect of learning to trust most contributes to the suboptimal learning we observe.

In the ultimatum game, future research should establish if there is a way to nudge people into more or less prosocial thinking when learning the different responder cultures. For example, participants could, through psychological framing, be made to temporarily harbor relatively selfish (prosocial) preferences; the question becomes whether this will diminish (enhance) the negative effects on learning we observe in our current sample. In other words: would altered social preferences lead to altered learning? And if so, would this also diminish the differences in beliefs we observed between social and non-social contexts? Another interesting question that future research can address is the degree to which the altered beliefs in the social condition depend on the act of learning. Do participants need to experience the results of their own actions in order to form these false beliefs, or would a similar effect occur if they were to simply observe someone else going through the motions?

Another interesting question that this thesis raises regards how performance in one of these tasks predict performance in the others. For example, do people who are very mild attackers also exhibit high inequity aversion and high levels of trust? Do people who are aggressive defenders have low levels of trust? Contingent trust? This will be a difficult question to address, but understanding how behavior in these different games relate will contribute to our understanding of how the norms contained within each game interact with one another.

Ultimately, each chapter acts as a building block contributing a different perspective to the study of human sociality. Using economic games, computational models based on the principle of utility, and model-based neuroimaging (Chapters 2 and 4), my research contributes to the scientific endeavor working to crack the "elaborate and secret code that is written nowhere, known by none, and understood by all" (Sapir, 1927, p.137).

References

Balliet, D., & Van Lange, P. A. M. (2013). Trust, conflict, and cooperation: A meta-analysis. *Psychological Bulletin*, *139*(5), 1090–1112. https://doi.org/10.1037/a0030939

Balodis, I. M., Kober, H., Worhunsky, P. D., Stevens, M. C., Pearlson, G. D., & Potenza, M. N. (2012). Diminished frontostriatal activity during processing of monetary rewards and losses in pathological gambling. *Biological Psychiatry*, *71*(8), 749–757.

Baumgartner, T., Fischbacher, U., Feierabend, A., Lutz, K., & Fehr, E. (2009). The Neural Circuitry of a Broken Promise. *Neuron*. https://doi.org/10.1016/j.neuron.2009.11.017

Behrens, T. E. J., Hunt, L. T., & Rushworth, M. F. S. (2009). The Computation of Social Behavior. *Science*, *324*(5931), 1160–1164. https://doi.org/10.1126/science.1169694

Behrens, T. E. J., Hunt, L. T., Woolrich, M. W., & Rushworth, M. F. S. (2008). Associative learning of social value. *Nature*, *456*(7219), 245–249. https://doi.org/10.1038/nature07538

Behrens, T. E. J., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. S. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience*, *10*(9), 1214–1221. https://doi.org/10.1038/nn1954

Bennett, C., Miller, M., & Wolford, G. (2009). Neural correlates of interspecies perspective taking in the post-mortem Atlantic Salmon: an argument for multiple comparisons correction. *NeuroImage*. https://doi.org/10.1016/s1053-8119(09)71202-9

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

Bicchieri, C. (2005). *The grammar of society: The nature and dynamics of social norms*. *The Grammar of Society: The Nature and Dynamics of Social Norms*. https://doi.org/10.1017/CBO9780511616037

Blake, P. R., McAuliffe, K., Corbit, J., Callaghan, T. C., Barry, O., Bowie, A., … Warneken, F. (2015). The ontogeny of fairness in seven societies. *Nature*, *528*(7581), 258–261. https://doi.org/10.1038/nature15703

Bohnet, I., Greig, F., Herrmann, B., & Zeckhauser, R. (2008). Betrayal Aversion: Evidence from Brazil, China, Oman, Switzerland, Turkey, and the United States. *American Economic Review*, *98*(1), 294–310. https://doi.org/10.1257/aer.98.1.294

Cacioppo, J. T., & Hawkley, L. C. (2009). Perceived social isolation and cognition. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2009.06.005

Camerer, C. (2003). Behavioural studies of strategic thinking in games. *Trends in Cognitive Sciences*, *7*(5), 225–231. https://doi.org/10.1016/S1364-6613(03)00094-9

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

Camerer, C., Ho, T. H., & Chong, J. K. (2002). Sophisticated Experience-Weighted Attraction Learning and Strategic Teaching in Repeated Games. *Journal of Economic Theory*, *104*(1), 137–188. https://doi.org/10.1006/jeth.2002.2927

Camerer, C., Ho, T. H., & Chong, J. K. (2004). A cognitive hierarchy model of games. *Quarterly Journal of Economics*, *119*(3), 861–898. https://doi.org/10.1162/0033553041502225

Camerer, C., Issacharoff, S. I., Loewenstein, G., Amerer, C. O. C., Ssacharoff, S. A. I., & Oewenstein, G. E. L. (2003). Regulation for conservatives: behavioral economics and the case for "asymmetric paternalism." *University of Pennsylvania Law Review*, (1211), 1211–1254.

Carter, J. R., & Anderton, C. H. (2001). An experimental test of a predator-prey model of appropriation. *Journal of Economic Behavior and Organization*, *45*(1), 83–97.

Charpentier, C. J., Aylward, J., Roiser, J. P., & Robinson, O. J. (2017). Enhanced Risk Aversion, But Not Loss Aversion, in Unmedicated Pathological Anxiety. *Biological Psychiatry*, *81*(12), 1014–1022.

Chen, S., & Bao, F. S. (2015). Linking body size and energetics with predation strategies: A game theoretic modeling framework. *Ecological Modelling*. https://doi.org/10.1016/j.ecolmodel.2015.07.033

Chowdhury, S. M., Jeon, J. Y., & Ramalingam, A. (2018). Property rights and loss aversion in contests. *Economic Inquiry*, *56*(3), 1492–1511. https://doi.org/10.1111/ecin.12505

Coricelli, G., & Nagel, R. (2009). Neural correlates of depth of strategic reasoning in medial prefrontal cortex. *Proceedings of the National Academy of Sciences*, *106*(23), 9163–9168. https://doi.org/10.1073/pnas.0807721106

Daw, N. D. (2011). Trial-by-trial data analysis using computational models. *Decision Making, Affect, and Learning: Attention and Performance XXIII*, 1–26. https://doi.org/10.1093/acprof:oso/9780199600434.003.0001

Dawkins, R., & Krebs, J. R. (1979). Arms races between and within species. *Proceedings of the Royal Society of London - Biological Sciences*. https://doi.org/10.1098/rspb.1979.0081

De Dreu, C. K. W., Giacomantonio, M., Giffin, M. R., & Vecchiato, G. (2019). Psychological constraints on aggressive predation in economic contests. *Journal of Experimental Psychology: General*, *148*(10), 1767–1781. https://doi.org/10.1037/xge0000531

De Dreu, C. K. W., & Gross, J. (2019). Revisiting the form and function of conflict: Neurobiological, psychological, and cultural mechanisms for attack and defense within and between groups. *Behavioral and Brain Sciences*, *42*(2019), e116. https://

doi.org/10.1017/S0140525X18002170

De Dreu, C. K. W., Gross, J., Méder, Z., Giffin, M., Prochazkova, E., Krikeb, J., & Columbus, S. (2016). In-group defense, out-group aggression, and coordination failures in intergroup conflict. *Proceedings of the National Academy of Sciences*, 201605115. https://doi.org/10.1073/pnas.1605115113

De Dreu, C. K. W., Kret, M. E., & Sligte, I. G. (2016). Modulating prefrontal control in humans reveals distinct pathways to competitive success and collective waste. *Social Cognitive and Affective Neuroscience*, *11*(8), 1236–1244.

De Dreu, C. K. W., Nijstad, B. A., Bechtoldt, M. N., & Baas, M. (2011). Group Creativity and Innovation: A Motivated Information Processing Perspective. *Psychology of Aesthetics, Creativity, and the Arts*. https://doi.org/10.1037/a0017986

De Dreu, C. K. W., Scholte, H. S., van Winden, F. a a M., & Ridderinkhof, K. R. (2015). Oxytocin tempers calculated greed but not impulsive defense in predator-prey contests. *Social Cognitive and Affective Neuroscience*, *10*(5), 721–728.

Delgado, M. R., Schotter, A., Ozbay, E. Y., & Phelps, E. A. (2008). Understanding Overbidding: Using the Neural Circuitry of Reward to Design Economic Auctions. *Science*, *321*(5897), 1849–1852.

Dimand, M.-A., & Dimand, R. (1996). *History Of Game Theory, Volume 1 : From the Beginnings to 1945. Routledge studies in the history of economics*.

Dorris, M. C., & Glimcher, P. W. (2004). Activity in posterior parietal cortex is correlated with the relative subjective desirability of action. *Neuron*, *44*(2), 365–378. https://doi.org/10.1016/j.neuron.2004.09.009

Ebitz, R. B., & Hayden, B. Y. (2016). Dorsal anterior cingulate: A Rorschach test for cognitive neuroscience. *Nature Neuroscience*, *19*(10), 1278–1279. https://doi.org/10.1038/nn.4387

Eklund, A., Nichols, T. E., & Knutsson, H. (2016). Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *Proceedings of the National Academy of Sciences*, *113*(33), 201602413.

Engelmann, J. B., Schmid, B., De Dreu, C. K. W., Chumbley, J., & Fehr, E. (2019). On the psychology and economics of antisocial personality. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(26), 12781–12786. https://doi.org/10.1073/pnas.1820133116

Erev, I., & Roth, A. E. (1998). Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria. *American Economic Review*, *88*(4), 848–881. https://doi.org/10.2307/117009

Fehr, E., & Camerer, C. F. (2007). Social neuroeconomics: the neural circuitry of social preferences. *Trends in Cognitive Sciences*, *11*(10), 419–427. https://doi.org/10.1016/j.tics.2007.09.002

Fehr, E., & Schmidt, K. M. (1999). A Theory of Fairness, Competition, and Cooperation. *The Quarterly Journal of Economics*, *114*(3), 817–868. https://doi.

org/10.1162/003355399556151

Fehr, E., & Schurtenberger, I. (2018). Normative foundations of human cooperation. *Nature Human Behaviour*, *2*(7), 458–468. https://doi.org/10.1038/s41562-018-0385-5

Friston, K. J., Parr, T., Zeidman, P., Razi, A., Flandin, G., Daunizeau, J., … Lambert, C. (2020). Tracking and tracing in the UK: a dynamic causal modelling study. Retrieved from http://arxiv.org/abs/2005.07994

Georgescu-Roegen, N. (1968). Utility. In D. L. Sills (Ed.), *International Encyclopedia of the Social Sciences* (15th ed., pp. 236–267). New York: Macmillan.

Gershman, S. J., & Uchida, N. (2019). Believing in dopamine. *Nature Reviews Neuroscience*, *20*(11), 703–714. https://doi.org/10.1038/s41583-019-0220-7

Grossman, H. I., & Kim, M. (1996). Predation and accumulation. *Journal of Economic Growth*, *1*(3), 333–350.

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

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

Heap, S. P. H., & Zizzo, D. J. (2009). The value of groups. *American Economic Review*, *99*(1), 295–323. https://doi.org/10.1257/aer.99.1.295

Henrich, J. (2016). *The secret of our success: How culture is driving human evolution, domesticating our species, and making us smarter.* Princeton, NJ, US: Princeton University Press.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., … Tracer, D. (2005). "Economic man" in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. *Behavioral and Brain Sciences*, *28*(6), 795–815. https://doi.org/10.1017/S0140525X05000142

Henrich, J., Ensminger, J., McElreath, R., Barr, A., Barrett, C., Bolyanatz, A., … Ziker, J. (2010). Markets, religion, community size, and the evolution of fairness and punishment. *Science*, *327*(5972), 1480–1484. https://doi.org/10.1126/science.1182238

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*. https://doi.org/10.1017/S0140525X0999152X

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

Jolly, E., & Chang, L. J. (2019). The Flatland Fallacy: Moving Beyond Low–Dimensional Thinking. *Topics in Cognitive Science*, *11*(2), 433–454. https://doi.org/10.1111/tops.12404

Kahneman, D., & Tversky, A. (1984). Choices. Values. Frames Kahnemann. *American Psychologist*, *39*(4), 341–350. https://doi.org/10.1037/0003-066X.39.4.341

Kluwer, E. S., Heesink, J. A. M., & Vliert, E. Van De. (1997). The Marital Dynamics of Conflict over the Division of Labor. *Journal of Marriage and the Family*. https://doi.org/10.2307/353951

Kolling, N., Wittmann, M. K., Behrens, T. E. J., Boorman, E. D., Mars, R. B., & Rushworth, M. F. S. (2016). Value, search, persistence and model updating in anterior cingulate cortex. *Nature Neuroscience*, *19*(10), 1280–1285.

Ku, G., Malhotra, D., & Murnighan, J. K. (2005). Towards a competitive arousal model of decision-making: A study of auction fever in live and Internet auctions. *Organizational Behavior and Human Decision Processes*, *96*(2), 89–103. https://doi.org/10.1016/j.obhdp.2004.10.001

Larson, P. B. (2010). Zermelo 1913. In *Ernst Zermelo - Collected Works/Gesammelte Werke* (pp. 260–273). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-79384-7_9

Levy, D. J., & Glimcher, P. W. (2012). The root of all value: a neural common currency for choice. *Current Opinion in Neurobiology*, *22*(6), 1027–1038. https://doi.org/10.1016/j.conb.2012.06.001

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., & Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature*. https://doi.org/10.1038/35084005

Lynch, P. (2008). The origins of computer weather prediction and climate modeling. *Journal of Computational Physics*. https://doi.org/10.1016/j.jcp.2007.02.034

McNamee, D., Rangel, A., & O'Doherty, J. P. (2013). Category-dependent and category-independent goal-value codes in human ventromedial prefrontal cortex. *Nature Neuroscience*, *16*(4), 479–485.

Meder, D., Haagensen, B. N., Hulme, O., Morville, T., Gelskov, S., Herz, D. M., … Siebner, H. R. (2016). Tuning the Brake While Raising the Stake: Network Dynamics during Sequential Decision-Making. *Journal of Neuroscience*, *36*(19), 5417–5426.

Metereau, E., & Dreher, J. C. (2015). The medial orbitofrontal cortex encodes a general unsigned value signal during anticipation of both appetitive and aversive events. *Cortex*, *63*, 42–54.

O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., & Dolan, R. J. (2004). Dissociable Roles of Ventral and Dorsal Striatum in Instrumental Conditioning. *Science*. https://doi.org/10.1126/science.1094285

Olsson, A., FeldmanHall, O., Haaker, J., & Hensler, T. (2018). Social regulation of survival circuits through learning. *Current Opinion in Behavioral Sciences*, *24*, 161–167. https://doi.org/10.31234/osf.io/r69e7

Ortner, S. B. (1998). Thick resistance: Death and the cultural construction of agency in Himalayan mountaineering. *Representations*. https://doi.org/10.2307/2928818

Ostrom, E. (1998). A Behavioral Approach to the Rational Choice Theory of Collective Action: Presidential Address, American Political Science Association, 1997. *American Political Science Review*, *92*(01), 1–22. https://doi.org/10.2307/2585925

Palminteri, S., & Pessiglione, M. (2017). Opponent brain systems for reward and punishment learning: Causal evidence from drug and lesion studies in humans. *Decision Neuroscience: An Integrative Perspective*, 291–303. https://doi.org/10.1016/B978-0-12-805308-9.00023-3

Palminteri, S., Wyart, V., & Koechlin, E. (2017). The Importance of Falsification in Computational Cognitive Modeling. *Trends in Cognitive Sciences*, *21*(6), 425–433. https://doi.org/10.1016/j.tics.2017.03.011

Perel, E. (2017). *The state of affairs : rethinking infidelity*.

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

Poldrack, R. A. (2011). Inferring mental states from neuroimaging data: From reverse inference to large-scale decoding. *Neuron*, *72*(5), 692–697. https://doi.org/10.1016/j.neuron.2011.11.001

Preuschoff, K., & Bossaerts, P. (2007). Adding Prediction Risk to the Theory of Reward Learning. *Annals of the New York Academy of Sciences*, *1104*(1), 135–146. https://doi.org/10.1196/annals.1390.005

Preuschoff, K., Quartz, S. R., & Bossaerts, P. (2008). Human Insula Activation Reflects Risk Prediction Errors As Well As Risk. *Journal of Neuroscience*, *28*(11), 2745–2752.

Prochazkova, E., Prochazkova, L., Giffin, M. R., Scholte, H. S., De Dreu, C. K. W., & Kret, M. E. (2018). Pupil mimicry promotes trust through the theory-of-mind network. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(31), E7265–E7274. https://doi.org/10.1073/pnas.1803916115

Romano, A., Balliet, D., & Wu, J. (2017). Unbounded indirect reciprocity: Is reputation-based cooperation bounded by group membership? *Journal of Experimental Social Psychology*. https://doi.org/10.1016/j.jesp.2017.02.008

Romano, A., Balliet, D., Yamagishi, T., & Liu, J. H. (2017). Parochial trust and cooperation across 17 societies. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1712921114

Rousseau, J. J., & Cole, G. D. H. (1993). *The Social Contract, and Discourses*. (J. H. Brumfitt & J. C. Hall, Eds.). New York: Alfred A. Knopf.

Rudorf, S., Preuschoff, K., & Weber, B. (2012). Neural Correlates of Anticipation Risk Reflect Risk Preferences. *The Journal of Neuroscience*, *32*(47), 16683–16692.

Rutledge, R. B., Dean, M., Caplin, A., & Glimcher, P. W. (2010). Testing the reward prediction error hypothesis with an axiomatic model. *Journal of Neuroscience*, *30*(40), 13525–13536. https://doi.org/10.1523/JNEUROSCI.1747-10.2010

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.

Sapir, E. (1927). The unconscious patterning of behavior in society. In *The unconscious: A symposium.* (pp. 114–142). New York: Alfred A. Knopf. https://doi.org/10.1037/13401-006

Schultz, W. (2010). Dopamine signals for reward value and risk: Basic and recent data. *Behavioral and Brain Functions*, *6*, 1–9. https://doi.org/10.1186/1744-9081-6-24

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*(5306), 1593–1599. https://doi.org/10.1126/science.275.5306.1593

Schwert, G. W. (2000). Hostility in takeovers: In the eyes of the beholder? *Journal of Finance*. https://doi.org/10.1111/0022-1082.00301

Silver, N. (2012). *The signal and the noise : why so many predictions fail--but some don't*. New York: Penguin Press.

Sutton, R., & Barto, A. (2018). Reinforcment Learning. *MIT Press*.

Tomasello, M. (2019). *Becoming human: A theory of ontogeny*. *Belknqp*.

Van Overwalle, F. (2009). Social cognition and the brain: A meta-analysis. *Human Brain Mapping*, *30*(3), 829–858. https://doi.org/10.1002/hbm.20547

Warren, D. E., Sutterer, M. J., Bruss, J., Abel, T. J., Jones, A., Kawasaki, H., … Tranel, D. (2017). Surgically disconnected temporal pole exhibits resting functional connectivity with remote brain regions. *Biorxiv*. https://doi.org/10.1101/127571

Wittmann, M. K., Kolling, N., Faber, N. S., Scholl, J., Nelissen, N., & Rushworth, M. F. S. (2016). Self-Other Mergence in the Frontal Cortex during Cooperation and Competition. *Neuron*, *91*(2), 482–493. https://doi.org/10.1016/j.neuron.2016.06.022

Xue, G., Lu, Z., Levin, I. P., Weller, J. A., Li, X., & Bechara, A. (2009). Functional dissociations of risk and reward processing in the medial prefrontal cortex. *Cerebral Cortex*, *19*(5), 1019–1027.

Zaki, J. (2014). Empathy: A motivated account. *Psychological Bulletin*, *140*(6), 1608–1647. Retrieved from http://psycnet.apa.org/journals/bul/140/6/1608/

Zaki, J., & Mitchell, J. P. (2013). Intuitive Prosociality. *Current Directions in Psychological Science*. https://doi.org/10.1177/0963721413492764

Zhu, L., Mathewson, K. E., & Hsu, M. (2012). Dissociable neural representations of reinforcement and belief prediction errors underlie strategic learning. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1116783109

2

Chapter 2

Neurocognitive Underpinnings of Aggressive Predation in Economic Contests

Michael Rojek-Giffin,[1,2,*] Mael Lebreton,[3,4] H. Steven Scholte,[5]
Frans van Winden,[6] K. Richard Ridderinkhof,[5] & Carsten K.W. De Dreu[1,2,5,*]

[1] Institute of Psychology, Leiden University, Leiden, The Netherlands, 2300 RB
[2] Institute for Brain and Cognition, Leiden University, Leiden, The Netherlands, 2300 RB
[3] Laboratory for Behavioral Neurology and Imaging of Cognition, Department of Basic Neuroscience, University of Geneva, Switzerland
[4] Swiss Center for Affective Sciences, University of Geneva, Switzerland
[5] Department of Psychology, University of Amsterdam, Amsterdam, The Netherlands, 1012 WX
[6] Center for Experimental Economics and Political Decision Making, University of Amsterdam, Amsterdam, The Netherlands, 1012 WX
[*] Corresponding author: Leiden Institute of Psychology, Wassenaarseweg 52, 2300 RB, Leiden, The Netherlands: m.r.giffin@fsw.leidenuniv.nl; +31 71 527 6684
or c.k.w.de.dreu@fsw.leidenuniv.nl; +31 71 527 3706

Summary

Competitions are part and parcel of daily life and require people to invest time and energy to gain advantage over others, and to avoid (the risk of) falling behind. Whereas the behavioral mechanisms underlying competition are well-documented, its neurocognitive underpinnings remain poorly understood. We addressed this using neuroimaging and computational modeling of individual investment decisions aimed at exploiting one's counterpart ("attack") or at protecting against exploitation by one's counterpart ("defense"). Analyses revealed that during attack relative to defense (I) individuals invest less and are less successful; (II) computations of expected reward are strategically more sophisticated (reasoning level $k = 4$; versus $k = 3$ during defense); (III) ventral striatum activity tracks reward prediction errors; (IV) risk prediction errors were not correlated with neural activity in either ROI- or whole-brain analyses; and (V) successful exploitation correlated with neural activity in the bilateral ventral striatum, left orbitofrontal cortex, left anterior insula, left temporoparietal junction, and lateral occipital cortex. We conclude that in economic contests, coming out ahead (versus not falling behind) involves sophisticated strategic reasoning that engages both reward and value computation areas and areas associated with theory of mind.

Key Words: Competition | K-level Reasoning | Theory of Mind | Reward Prediction | Risk

Author Note

Financial support was provided by a seed-grant from the Amsterdam Brain and Cognition Priority Area to CKWDD, FVW and RR, an SNF Ambizione Grant (PZ00P3_174127) to ML, and the Spinoza Award from the Netherlands Science Foundation (NWO SPI-57-242) to CKWDD. CKWDD, RR and FVW conceived of the study and designed the behavioral experiment. HSS implemented and coordinated neuroimaging. MRG and ML contributed the computational model and analyzed the data. MRG and CKWDD wrote the paper and incorporated co-author revisions. The authors note that this study capitalizes on computational models that were conceived after data collection and that were not part of our original set of predictions and analysis plans. No conflict of interest declared.

Introduction

In his *principles of political economy*, John Stuart Mill, (1859) observed that "a great proportion of all efforts … [are] spent by mankind in injuring one another, or in protecting against injury." Such appetite for "injuring others" and to defend against being injured has recently been documented in economic contest experiments in which individuals invest to obtain a reward at a cost to their competitor (henceforth attack), or to avoid losing their resources to their antagonist (henceforth defense; Carter & Anderton, 2001; Chen & Bao, 2015; Chowdhury, Jeon, & Ramalingam, 2018; De Dreu & Gross, 2019; De Dreu, Kret, & Sligte, 2016; De Dreu, Scholte, van Winden, & Ridderinkhof, 2015; Grossman & Kim, 1996; Wittmann et al., 2016; Zhu, Mathewson, & Hsu, 2012). These experiments showed that humans invest in injuring others through attacks and in protecting against injuring through defense, that investments in attack are typically less frequent and forceful than investments in defense, and that attack decisions disproportionally often fail and defenders relatively often survive (with ≈ 30% victories against ≈ 70% survivals) (for a review see e.g., De Dreu & Gross, 2019).

Resonating with the idea that competition can be costly, participants during such attacker-defender contests typically waste about 40% of their wealth in fighting each other (De Dreu & Gross, 2019). Yet why people invest in attack and defense remains poorly understood. In fact, investing in injuring others, and in protecting against injury, may reflect an array of subjective "desires" (Charpentier, Aylward, Roiser, & Robinson, 2017; Delgado, Schotter, Ozbay, & Phelps, 2008; Dorris & Glimcher, 2004). Perhaps humans invest in attack and defense to maximize their personal earnings, as is typically assumed in standard economic theory (e.g. Ostrom, 1998). Relatedly, individuals may invest in attack and defense because of "competitive arousal" and rivalry (Delgado et al., 2008; Ku, Malhotra, & Murnighan, 2005). Finally, investment in attack and defense may be driven by a desire to minimize risk and uncertainty (Delgado et al., 2008; Kahneman & Tversky, 1984). Indeed, decision-making in competitive contests is inherently risky – investments are typically wasted and may result in no return (among attackers), in wasted resources (when attacks were unexpectedly shallow and one thus over-invested in defense), or in costly defeat (when attacks were unexpectedly tough). Humans factor in such risks when making decisions and are typically risk-averse (Kuhnen & Knutson, 2005; Loewenstein, Hsee, Weber, & Welch, 2001; Tobler, O'Doherty, Dolan, & Schultz, 2006).

Humans may hold conflicting desires when investing in attack and defense, and may need to balance between maximizing reward and minimizing risk. What individuals aim for and how possibly conflicting desires are regulated is difficult to infer from behavioral decision-making alone. To illustrate, consider a two-player contest in which one participant can invest in attack and the other participant in defense. When the attacker invests more than its defender, attackers obtain all what the defender did not invest and the defender would be left with 0. If, attackers invests equal or less than their defender, both sides earn their non-invested resources (Carter & Anderton, 2001; Chowdhury et al., 2018; De Dreu & Gross, 2019; De Dreu, Gross, et al., 2016; De Dreu, Kret, et al., 2016; De Dreu et al., 2015; Grossman & Kim, 1996)[1]. It follows that investments can increase attacker earnings and their competitive success, and can prevent defenders from losing their remaining endowment to their attacker. At the same time, however, not investing resources eliminates the attacker's uncertainty about earnings from the contest, alongside the possibility of losing money. Defenders, in contrasts, reduce such uncertainty and possibility of losing the contest by investing resources (Chowdhury et al., 2018).

We solved this problem of inference using a two-pronged approach inspired by recent work in cognitive neuroscience on learning from reward and risk prediction (Olsson, FeldmanHall, Haaker, & Hensler, 2018; Palminteri, Wyart, & Koechlin, 2017; Preuschoff, Quartz, & Bossaerts, 2008). First, from investments in attacker-defender contests we computed, using a k-level reasoning approach, estimates of expected reward and expected risk (Botvinick, Niv, & Barto, 2009; Camerer, Ho, & Chong, 2004; Harsanyi, 1967; Nagel, 2016; Ribas-Fernandes et al., 2011; Stahl & Wilson, 1995; Zhu et al., 2012). The computational approach incorporates the intuition that the formation of expectations and beliefs in strategic interactions are recursive (i.e., [1] I think that [2] you think that [3] I think that [4]…) and can be more or less sophisticated (i.e., the number of recursions *k*). Using computational modeling and model comparison we estimated for each investment in attack and defense the expected reward and risk, and concomitant reward and risk prediction errors. Our modeling thus defines (expected) reward as the (expected) monetary payoff from investment in attack and defense (e.g., Zhu et al., 2012), and (expected) risk as the (expected) variance of the reward prediction error (Preuschoff et al., 2008).

Second, and next to an exploratory whole-brain analysis potentially revealing currently unknown cues about the neural foundations of exploitation and protection, we linked prediction errors to *a priori* defined regions of interest—the Ventral Striatum and the Amygdala. We chose the ventral striatum because it has been extensively linked to reward processing and competitive success (*viz*. reward maximization; Balodis et al., 2012; McNamee, Rangel, & O'Doherty, 2013; Metereau & Dreher, 2015; Rudorf, Preuschoff, & Weber, 2012; Xue et al., 2009; Zhu et al., 2012). We chose the Amygdala because of its involvement in low-level affective processing of threat to resources (*viz.* risk minimization; Baumgartner, Heinrichs, Vonlanthen, Fischbacher, & Fehr, 2008;

1 The attack-defense contest belongs to a class of asymmetric conflict games in which one player competes to maximize personal gain and the counterpart competes to prevent exploitation (De Dreu & Gross, 2019; Dechenaux, Kovenock, & Sheremeta, 2015). Including in this class of asymmetric games are the Hide-and-Seek game (Bar-Hillel, 2015; Flood, 1972), the matching-pennies game (Goeree, Holt, & Palfrey, 2003), the inspection game (Nosenzo, Offerman, Sefton, & van der Veen, 2014), and the Best-shot/Weakest-link game (Chowdhury & Topolyan, 2016; Clark & Konrad, 2007). Across these games, humans invest to maximize wealth and/or to minimize risk of losing.

Choi & Kim, 2010; De Dreu et al., 2015; Delgado et al., 2008; Nelson & Trainor, 2007; Phelps & LeDoux, 2005).

Materials and Methods

Participants and Ethics

Male participants (M = 25.31 years; N = 27) were recruited via an on-line recruiting system for participating in a neuro-imaging study on human decision-making. Exclusion criteria were significant neurological or psychiatric history, prescription-based medication, smoking more than five cigarettes per day, and drug or alcohol abuse.[2] Eligible participants were assigned to a session and instructed to refrain from smoking or drinking (except water) for 2 hours before the experiment that lasted approximately 1.5 hours. They received a show-up fee of €30 in addition to the earnings from decision making. The experiment involved no deception and was incentivized (see below), received ethics approval from the Psychology Ethics Committee of the University of Amsterdam, and complied with the guidelines from the American Psychological Association (6th edition). Participants provided written informed consent before the experiment and received a full debriefing afterwards.

Experimental Procedures

Experimental sessions were conducted between noon and 4PM and participants were tested individually (also see De Dreu et al., 2015). Upon arrival, participants were escorted to a private cubicle where they read and signed an informed consent form. Participants received a booklet with instructions for the Attacker-Defender Game (labeled Investment Task), containing several examples of investments and their consequences to both attacker (labeled Role A) and defender (labeled Role B), and several questions to probe understanding of the game structure and decision consequences. Neutral labeling was used throughout.

Upon finishing the instructions for the contest, the experimenter prepared the participant for neuro-imaging. During the *f*MRI session, participants completed 6 functional runs, each consisting of a 20 trial block played as either attacker or defender. Participants thus alternated between the role of attacker and defender every 20 trials, with the starting order counter-balanced across participants. Importantly, we used a random-partner matching one-shot protocol, eliminating reputation concerns (Zhu et al., 2012). In each session, participants made 60 investments as attacker, and 60

2 The sample was the same as used in De Dreu et al. (2015), which used a cross-over design to examine the behavioral and neural effects of oxytocin (versus placebo) administration. Here we only analyze investments made under placebo. Moreover, our earlier report only considered trials in which participant decisions affected themselves only, and did not include those decision trials in which decisions also affected two other individuals within their group. Here we include also those previously unanalyzed trials. Because this manipulation revealed no differences, we collapsed across these two conditions. In short, the current study shares 25% of its analyzed data with the previous one, asks a different research question and uses distinctly different analytic techniques.

as defender. For each investment trial, they received a prompt, randomly generated between 0 (indicating no investment) and 10 (indicating investment of the entire endowment) and used a button-press to adjust the given number up or down to indicate their desired investment. The duration of the selection period was self-paced, and had an average length of 4.27 seconds (SD = 3.43 seconds) (see Figure 1). After selecting their investments, participants waited an average of 6.08 seconds (SD = 2.22 seconds), at which point they received feedback about their counterpart's investment, and were shown the respective payoffs to oneself and the other (who was randomly chosen on each trial from a pool of 150 attacker [defender] investments; for further detail see De Dreu et al., 2019, 2015). At the end of the experiment participants received their participation fee and earnings by bank transfer (range €0 – €8, with *M* = €5 for non-scanner participants, and €0 – €33, with *M* = €19 for scanner participants). Accordingly, participant pay was private and conditioned on their performance.

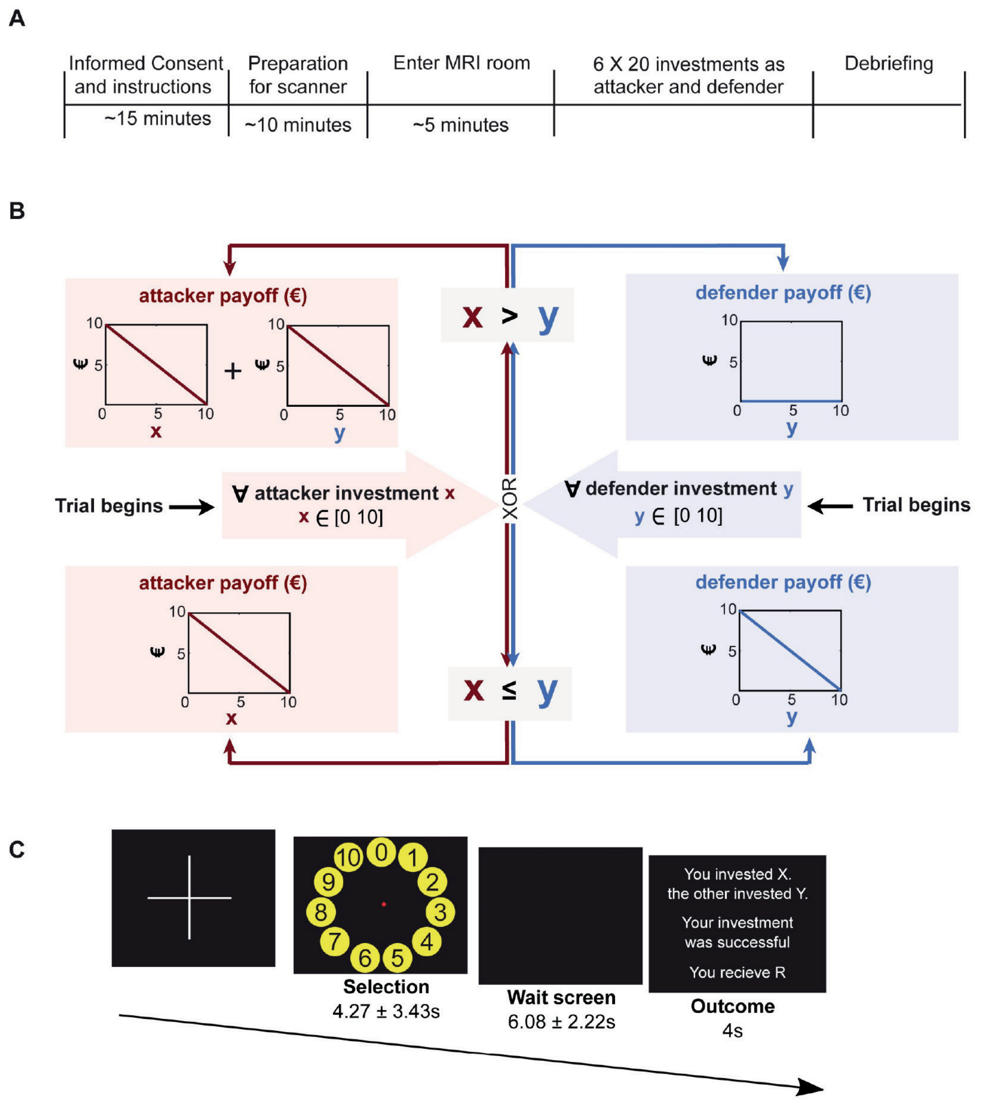

Figure 1. Experimental design. (A) Timeline of the entire experiment. (B) The Attacker Defender contest: on each trial, both attackers and defender begin with a 10€ endowment with which to invest in the contest. Investments are non-recoverable, yet if the defender invests equal or more than the attacker (bottom), both attacker and defender keep their remaining endowments (i.e. whatever they did not invest in the contest). If the attacker invests more than the defender (top), the attacker receives their remaining endowment plus that of the defender, who receives nothing. (C) Trial break-down: for each trial, participants received a prompt, randomly generated between 0 (indicating no investment) and 10 (indicating investment of the entire endowment) and used a button-press to adjust the given number up or down to indicate their desired investment. The duration of the selection period was self-paced (M ± SD = 4.27 ± 3.43 seconds). After selecting their investments, participants waited an average of M ± SD = 6.08 ± 2.22 seconds and then received feedback about their counterpart's investment and the payoffs to oneself and to the counterpart. This completed one trial.

Attacker-Defender Contest

The Attacker-Defender Contest (Figure 1B) consists of two players: an attacker and a defender. Each player was endowed with €10 from which they could invest in the contest. Investments were always wasted but if the investments by the attacker (x) exceeded that by the defender (y), the attacker ($x > y$) the attacker obtains all of the defender's non-invested endowment ($e–y$). In this case, the attacker's total earning was $2e–x–y$, and the defender earned 0. If, in contrast, the defenders investment matched or exceeded that by the attacker ($y \geq x$), both defender and attacker earned what was left from their endowment ($e – y$, and $e – x$, respectively) (De Dreu et al., 2015; 2016ab; 2019).

The Attacker-Defender Contest has a contest success function $f = X^m/(X^m + Y^m)$, where f is the probability that the attacker wins, $m \to \infty$ for $X \neq Y$ and $f = 0$ if $Y = X$. Assuming rational selfish play and risk-neutrality, standard economic theory predicts that attackers and defenders use mixed strategies when investing. With e = 10€ per trial (as used in the current experiment), the mixed strategies for attack (with probability of investing x denoted by p(x)) and defense (with probability of investing y denoted by p(y)) define a unique Nash equilibrium where expected investments in attack are both lower (x = 2.62) than in defense (y = 3.38), and less frequent (probability of attack [defense] = 60% [90%]). However, when attacks are made they are expected to be more 'forceful' (4.36 versus 3.75 for defense).[3]

Modeling Investment Behavior with K-level Sophistication

To compute individual estimates of expected reward and concomitant reward and risk prediction errors, we adapted the cognitive-hierarchies framework developed in behavioral economics (Botvinick et al., 2009; Camerer et al., 2004; Nagel, 2016). The idea is that players hierarchically form beliefs about their opponents' behavior, up to a certain level of cognitive sophistication (k-level). A k-0 player invests randomly. At k = 1 the individual assumes that her opponent has k = 0 and finds an investment that maximizes her expected reward under this assumption. At k = 2 the individual assumes that her opponent has k = 1 and finds an investment that maximizes her own expected reward under the assumption that the opponent seeks to maximize his personal reward against a k-0 player. This recursion can, in theory, continue infinitely, yet in our computational modeling we limited $k \leq 5$. ***k-level 0.*** k-level 0 play each strategy with equal probability. We have: Specifically, when I_S represent a player's own investment (s stands for *self*) and I_O their representation of the other player's investment (o stands for *other*) we can formally express:

3 Specifically, the mixed-strategy equilibrium is computed as follows: Attack: $p(x=1) = 2/45$, $p(x) = p(x–1)[(12–x)/(10–x)]$ for $2 \leq x \leq 6$, $p(x=0) = 1–[p(x=1) +…+ p(x=6)] = 0.4$, and $p(x) = 0$ for $x \geq 7$; Defense: $p(y) = 1/(10–y)$ for $0 \leq y \leq 5$, $p(y=6) = 1 – [p(y=0) +…+ p(y=5)] = 0.15$, and $p(y) = 0$ for $y \geq 7$ (also see De Dreu et al., 2015).

k-level 0. k-level 0 play each strategy with equal probability. We have:

$$\forall\, h \in \{0, \ldots, 10\}, P(I_s = h) = \frac{1}{11} \tag{1}$$

k-level 1. k-level 1 expect their opponent to play as k-level 0, such that they expect:

$$\forall\, h \in \{0, \ldots, 10\}, P(I_o = h) = \frac{1}{11} \tag{2}$$

These expectations can be used to compute the probability of success S of a given investment h (P(S|h)) by the attacker A and defender D, respectively:

$$\forall\, h_A \in \{0, \ldots, 10\}, P(S|h_A) = \sum_{i=0}^{hA\text{-}1} P(I_o = i) \tag{3}$$

$$\forall\, h_D \in \{0, \ldots, 10\}, P(S|h_D) = \sum_{i=0}^{hD} P(I_o = i) \tag{4}$$

This can be used to compute an expected value, which in this case in the expected reward *ER* for any potential investment by the attacker and defender. We have, for the attacker:

$$ER_A(h_A) = [P(S|h_A) \times (10\text{- } E[h_D|h_D < h_A] + 10\text{-}h_A)] + [(1\text{–}(S|h_A)) \times (10\text{-}h_A)] \tag{5}$$

where the two square brackets represent cases where the investment is successful or unsuccessful, respectively, and $E[h_D|h_D < h_A]$ is the expected opponent's investment in case of success:

$$E[h_D|h_D < h_A] = \sum_{i=0}^{hA\text{-}1} i \times P(I_o = i) \tag{6}$$

For the defender we have, likewise:

$$ER_D(h_D) = [P(S|h_D) \times (10\text{-}h_D)] + [(1\text{-}P(S|h_D)) \times 0] \tag{7}$$

The expected reward also has an associated prediction error *PE*, which is simply the expected reward *ER* subtracted from the actual reward *R*

$$PE = R\text{ -}ER \tag{8}$$

These values also allow for the calculation of risk prediction *RP* and accompanying risk prediction errors PE_{Risk}. We defined risk prediction as the expected size-squared of the reward prediction error (Preuschoff et al., 2008). More specifically, risk prediction is defined as the sum across all the possible rewards (*R*) of $(R - ER)^2$, multiplied by the probability $P(R)$ that R is obtained. More formally:

$$\mathrm{RP} = \mathrm{E}[(\mathrm{R\text{-}ER})^2] = \sum_{\mathrm{R}} \mathrm{P(R)} \times (\mathrm{R\text{-}ER})^2 \tag{9}$$

Which means that the risk prediction error PE_{Risk} is the risk prediction RP subtracted from the actual size-squared of the reward prediction error:

$$\mathrm{PE_{Risk}} = (\mathrm{R\text{-}ER})^2\text{-}\mathrm{RP} \tag{10}$$

Following standard practices in the field, we assume that participants select the investment I_S that (soft-)maximizes their expected reward. This is modelled with a multinomial softmax function with free parameter β, which indexes the exploration/exploitation tradeoff (choice temperature):

$$\mathrm{P}(\mathrm{I_s} = \mathrm{h_i}) = \frac{\exp(\beta_1 \times \mathrm{EV}(\mathrm{h_i}))}{\sum_{\mathrm{i}=0}^{10} \exp(\beta_1 \times \mathrm{EV}(\mathrm{h_j}))} \tag{11}$$

This choice temperature defines the likelihood of investments I_S, i.e. the probability of observing investment I_S under the considered model and parameter values.

k-levels 2 → n.

For each k-level, k ≥ 2, the above procedure is iterated k-times, with k-level predictions of investments - needed to compute probabilities of success, expected rewards and choice probabilities - being generated by the softmax at the preceding level (see Figure 2). Hence, each k-level model has k free-parameters, which constitutes the choice temperature at each level β_k.

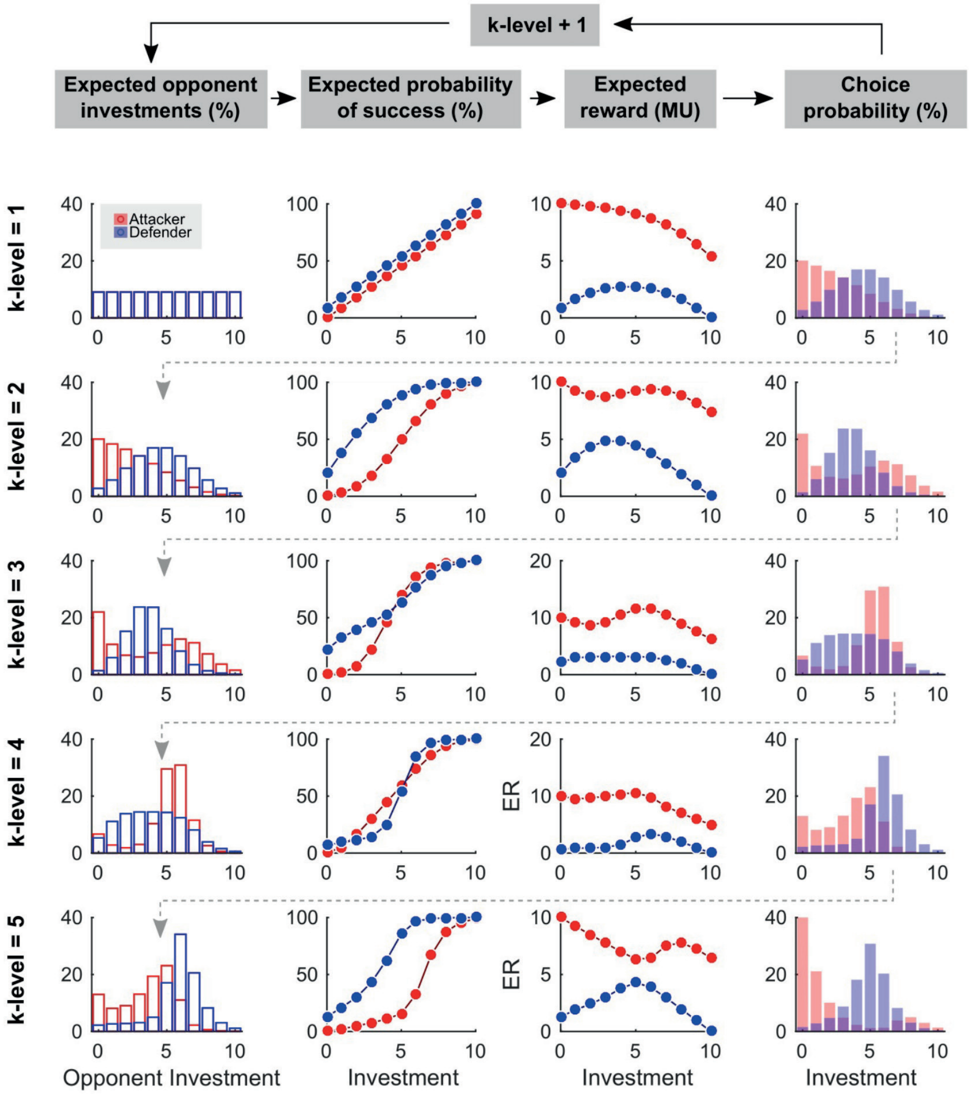

Figure 2. Computational framework. Players hierarchically form beliefs about their opponents' behavior, up to a certain level of cognitive sophistication (k-level) (column 1). The expected frequencies of the opponents investment are then used to calculated expected probability of success for each investment (column 2), which can then be used to calculated expected reward (column 3). Based on the expected reward, we calculate the frequency that a player should make each investment (column 4). A k-2 player (row 2) will assume that her opponent is k-1 and adjust her behavior accordingly, and so on. We developed computational models for hierarchies 1 up to 5.

Model fitting

For each model M, the parameters θ_M ($\theta_M = \{\beta_1, \beta_2, \ldots \beta_k\}$) were optimized by minimizing the negative logarithm of the posterior probability (*LPP*) over the free parameters:

$$\mathrm{LPP} = -\log\big(P(\theta_M|D, M)\big) \propto -\log\big(P(D|M, \theta_M)\big) - \log\big(P(\theta_M|M)\big) \tag{12}$$

Here, $P(D|M, \theta_M)$ is the likelihood of the data D (i.e. the observed choice) given the considered model M and parameter values θ_M, $P(\theta_M|M)$ and is the prior probability of the parameters. Following Daw (2011), the prior probability distributions were defined as a gamma distribution (gampdf(β,1.2,5)) for the choice temperature. This procedure was conducted using Matlab's fmincon function with different initialized starting points of the parameter space (i.e., 0<β<Infinite) (Palminteri, Khamassi, Joffily, & Coricelli, 2015). We computed the Laplace approximation to model evidence (ME). It measures the ability of each model to explain the experimental data by trading-off their goodness-of-fit and complexity. Defining $\hat{\theta}_M$ as the model parameters identified in the optimization procedure and *n* as the number of data-points (i.e. trials), ME was computed as follows (Where |H| is the determinant of the Hessian matrix):

$$\mathrm{ME} = \log\left(P\left(D|M, \hat{\theta}_M\right)\right) + \log\left(P\left(\hat{\theta}_M|M\right)\right) + \frac{df}{2}\log(2\pi) - \frac{1}{2}\log|H| \tag{13}$$

Bayesian Model Comparison.
To identify the model most likely to have generated a certain data set, ME was computed at the individual level for each model in the respective model-space, and fed to random-effects Bayesian Model Comparison using the mbb-vb-toolbox (http://mbb-team.github.io/VBA-toolbox/; Daunizeau, Adam, & Rigoux, 2014). This procedure estimates the expected frequencies (denoted PP) and the exceedance probability (denoted XP) for each model within a set of models, given the data gathered from all subjects. PP quantifies the posterior probability that the model generated the data for any randomly selected subject. XP quantifies the belief that the model is more likely than all the other models of the model-space. An XP > 95% for one model within a set is typically considered as significant evidence in favor of this model being the most likely.

Model identifiability.
To assess the reliability of our modelling approach, we performed model identifiability simulations (see Correa et al., 2018 for a similar approach). Choices from synthetic subjects were generated for each task and each model, by running our computational models, with model parameters sampled in their prior distribution: softmax temperature β were drawn from gamma distribution (random('Gamma',1.2,3)). For each model, we ran 10 simulations including 27 synthetic subjects (N=270), playing both attacker and

defender for 3 blocks of 20 trials. Model identifiability was assessed by running the Bayesian Model Comparison on the synthetic data.

MRI Data Acquisition, Preprocessing, and Data Analysis

Scanning was performed on a 3T Philips Achieva TX MRI scanner using a 32-channel head coil. Each participant played six blocks of the attacker-defender game in which functional data were acquired using a gradient-echo, echo-planar pulse sequence (TR=2000 ms, TE=27.63 ms, FA=76.18, 280 volumes, FOV=192^2 mm, matrix size=64^2, 38 ascending slices, slice thickness=3 mm, slice gap=0.3 mm) covering the whole brain. For each subject, we also recorded a 3DT1 recording (3D T1 TFE, TR=8.2 ms, TE=3.8 ms, FA=88, FOV=256^2 mm, matrix size=256^2, 160 slices, slice thickness=1 mm) as well as respiration, pulse oximetry signal, and breath rate. Stimuli were back-projected onto a screen that was viewed through a mirror attached to the head-coil.

Analyses were conducted with FSL (Oxford Centre for Functional MRI of the Brain (FMRIB) Software Library; www.fmrib.ox.ac.uk/fsl) and custom scripts written in Matlab (Mathworks, US). All fMRI data was pre-whitened, slice-time corrected, spatially smoothed with a 5mm FWHM gaussian kernel, motion corrected, and high-pass filtered. Functional images were registered to each subject's high resolution T1 scan and subsequently registered to MNI space.

Our primary goal was to determine if neural activity was modulated by the expected values and/or prediction errors from our reinforcement learning model. The entire *f*MRI analysis consisted of a 3-level analysis: level 1 was averaging within runs within subjects, level 2 was averaging across runs within subjects, and level 3 was testing for significance at the group level. We constructed 3 different general linear models (GLM's) to test for significant neural differences between attack and defense behavior as well as to see if attack and defense behavior correlated with our variables of interest. GLM-1 was meant to test for simple model-free differences between attacker and defender neural activity and consisted only of the selection and feedback epochs. GLM-2 was meant to determine if neural activity significantly correlated with investment magnitude during the selection time-phase and whether wins/losses significantly correlated with neural activity during feedback. To this end it consisted of the following regressors: selection, selection modulated by investment (orthogonalized with respect to selection), feedback, and feedback modulated by wins/losses (*z*-scored and orthogonalized with respect to feedback). GLM-3 was meant to determine whether any neural activity correlated with the parameters calculated from our K-Level model and contained the following regressors: selection, selection modulated by expected value (orthogonalized with respect to selection), selection delayed by 4 seconds in order to capture the delayed nature of risk prediction (Preuschoff et al., 2008), delayed selection modulated by risk prediction (orthogonalized with respect to delayed selection), feedback, feedback modulated by the prediction error (z-scored and orthogonalized with respect to feedback), and feedback modulated by the risk prediction error (z-scored orthogonalized with respect to feedback). To mitigate spurious results from asymmetric parameter value ranges (Lebreton, Bavard, Daunizeau, & Palminteri, 2019), each parametric regressor was z-scored within each role, meaning both attacker and defender parametric regressors had identical variance.

We checked for multicollinearity by calculating the variance inflation factors (VIF) for each regressor of interest (Mumford, Poline, & Poldrack, 2015), and found none to be problematic (all VIF's < 2.3). However, four subjects made identical investments on every trial, which resulted in rank deficient models (4 subjects for GLM-2 and GLM-3). Specifically, two individuals made the exact same investment on all attack decisions, one individual made the exact same investment on all defense decisions, and one individual made the exact same investment during attack and defense. These subjects had to be removed from the analysis. We tested for an interaction effect between role and each variable of interest by contrasting the relevant parameter estimates for attack and defense in a second level within-subject fixed-effects analysis. Finally, we tested for group level significance and corrected for multiple comparisons using FSL's FLAME 1 with the standard cluster forming threshold of $Z>3.1$ and clusters significant at $p = 0.05$. We ran additional control analyses with FSL's randomized threshold-free cluster enhancement (TFCE) (Smith & Nichols, 2009; Winkler, Ridgway, Webster, Smith, & Nichols, 2014), and results were virtually identical.

We also conducted analyses within an a priori selected anatomical ventral striatum (VS), and within an a priori selected anatomical amygdala ROI. Both masks were obtained from the meta-analytic tool Neurosynth (Yarkoni, Poldrack, Nichols, Van Essen, & Wager, 2011). We used the terms "ventral striatum" and "amygdala" in our search of Neurosynth, instead of using "reward" or "fear." Avoiding psychological constructs such as reward or fear reduced possible bias in our ROI's in favor of a particular psychological construct. For our ROI analyses, we took the average value across every voxel within each ROI for each subject within the contrast of interest (e.g. attacker-reward prediction error), and then tested for significance with a paired-sample t-test.

Results

Decision-Making

Earlier reports of the attacked-defender contest game analyzed investments in terms of the overall investment (range 0 – 10), the frequency of investment (all trials in which x or $y > 0$; range 0 – 60), and the force of investment (the amount invested on non-zero investment trials, range 1 – 10). For these measures we find, consistent with earlier work, that individuals invested less often in attack than in defense, $t(26) = -4.12$, $p = 0.0003$, invested in attack less overall, $t(26) = -8.56$, $p < 0.0001$, and invested less forcefully in attack than in defense, $t(26) = -7.81$, $p < 0.0001$ (Figure 3B). Although individuals earned more from attack (non-invested resources + spoils of winning) than defense trials (non-invested resources in case of survival), $t(26) = 43.91$, $p < 0.0001$, they were less successful during attack than defense trials, $t(26) = -7.22$, $p < 0.0001$: As defender they "survived" more often than that they "killed" as attacker (Figure 3C).

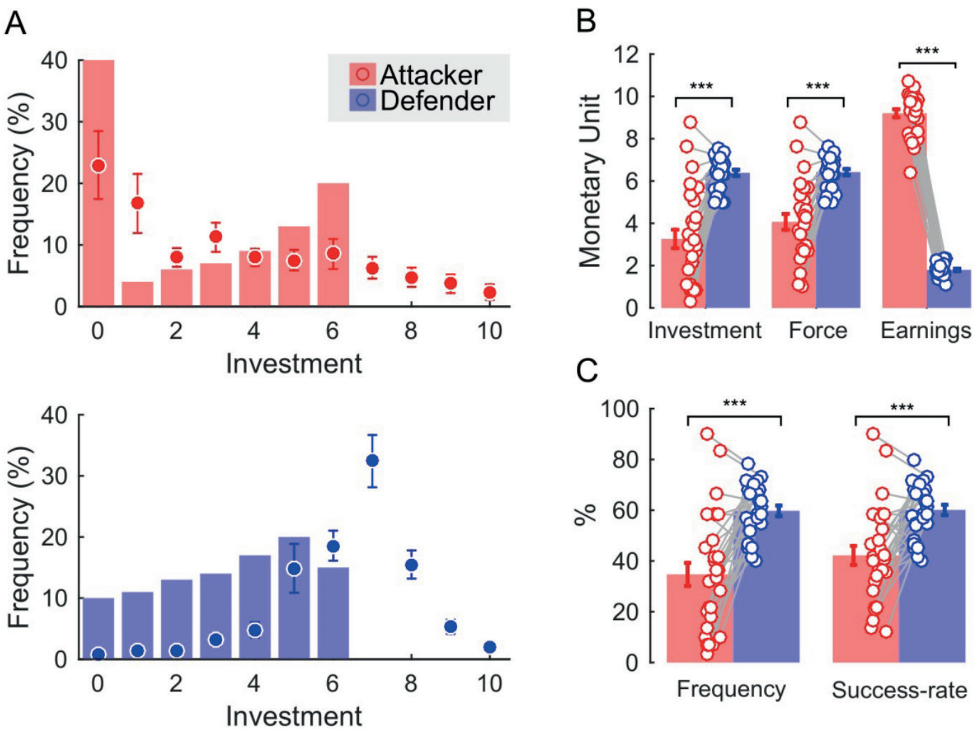

Figure 3. Behavioral results. (A) Nash equilibrium predictions (bars) plotted against empirical distribution of participants' investments (dots with error bars are Means ± 1 Standard Error) for attacker (top row, red) and defenders (bottom row, blue). (B) Attacker (red) and defender (blue) investments, force of investment, and mean earnings (shown are Means ± 1 Standard Error) (C) frequency of investment, and success-rate (shown are Means ± 1 Standard Error). Contrasts marked * (**) (***) are significant at p < 0.05 (0.01) (0.001).

In addition to the contrast between attack and defense, we examined investments in relation to predictions derived from standard economic theory that assumes rational self-interest and risk-neutrality. Relative to mixed-strategy equilibrium predictions (see *Materials and Method*), individuals invest more, and more forcefully in defense ($t(26) = 20.40$, $p < 0.0001$, and $t(26) = 18.467$, $p < 0.0001$, respectively), but not more, and not more forcefully in attack ($t(26) = 1.46$, $p = 0.157$, and $t(26) = -0.78$, $p = 0.441$, respectively) (Figure 3A). Still, however, both attack and defense returned less earnings than predicted by standard economic theory ($t(26) = -4.19$, $p = 0.00028$, and $t(26) = -40.56$, $p < 0.0001$), and the frequency of both attacks and defense exceeded expectations based on rational selfish play ($t(26) = 3.04$, $p = 0.0054$, and $t(26) = 30.26$, $p < 0.0001$, respectively). Conversely, success-rates for attacks (victories) and defense (survival) did not deviate from Nash equilibrium predictions ($t(26) = -0.25$, $p = 0.804$, and $t(26) = -0.98$, $p = 0.336$, respectively).

Neural Correlates of Attack and Defense.

To examine the neural foundations of decision-making during attack and defense, we performed whole-brain analyses on the selection phase (when subjects decided whether and how much to invest in attack or defense) and on the feedback phase (when subjected received information about their opponent's investment and the resulting outcomes to oneself). Whereas no significant differences between attacker and defender were observed during selection, whole-brain analyses did show significant attacker-defender contrasts for the feedback phase. Specifically, during feedback, participants exhibited higher BOLD response during attack relative to defense in a cluster within the left anterior insula and inferior frontal gyrus (Figure 4: MNI coordinates: x = -40, y = 10, z = 16, Z = 4.88, cluster size = 1657, $p = 0.0151$, *FWE*-whole brain).

Table 1: Regions exhibiting significant correlation between neural activity and win / loss feedback during attack.

	Peak					
Region	**x**	**y**	**z**	**Cluster size**	**Z-value**	**p (FWE-corr)**
Attacker Win/Loss						
VS/OFC/Insula/Thalamus	-8	4	-4	5329	4.27	<0.001
Lateral Occipital Cortex	-22	-74	-8	1686	4.75	0.002
Occipital Pole	8	-84	4	1603	4.45	0.002
TPJ/Lateral Occipital Cortex	-26	-84	46	1577	4.1	0.003

Note. All statistics are corrected for multiple comparison with FSL's FLAME 1.

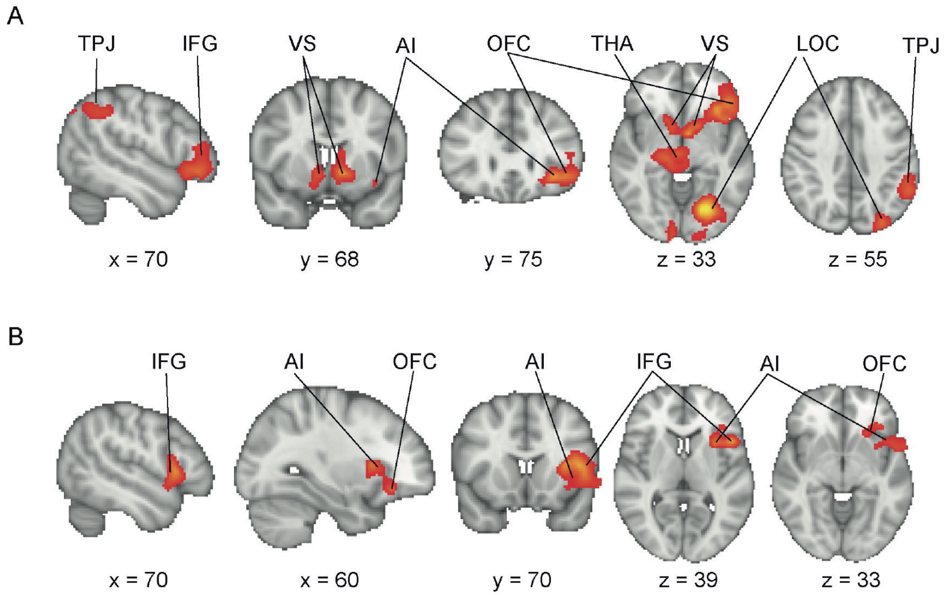

Figure 4. Brain-imaging Results. Whole brain analysis testing for attacker neural activity correlated to wins and losses (A), and feedback differences between attacker and defender (B). (A) Wins and losses as an attacker correlated with neural activity in the temporo-parietal junction (TPJ), inferior frontal gyrus (IFG), ventral striatum (VS), anterior insula (AI), thalamus (THA), and lateral occipital cortex (LOC). (B) Processing feedback as an attacker associated with more neural activation in the left inferior frontal gyrus (IFG), left anterior insula (AI), and left orbitofrontal cortex (OFC). All contrasts are FWE-corrected at $p < 0.05$ for the whole brain.

In a follow-up analysis we examined whether participant's exhibited a correlation between neural activity and investments (during decision-making) and outcome (win/loss) during feedback. As before, no significant correlations were found between neural activity and investments during attack or defense, nor did the correlation differ between the two roles. During feedback, however, neural activity during attack covaried with wins and losses in clusters that included the bilateral ventral striatum, left orbitofrontal cortex, left anterior insula, left temporoparietal junction, and lateral occipital cortex (Table 1, Figure 4B). Activity in these same areas also correlated with wins/losses more during attack than defense, but did not survive cluster-based multiple comparison correction (with $p < 0.05$, uncorrected). When participants processed feedback as defenders there were no clusters that significantly covaried with wins and losses.

Model-Based Analyses of Decision-Making and Neural Activity

As noted in the Methods, we captured the computations at hand in attack and defense behavior using the cognitive-hierarchies framework developed in behavioral economics (Botvinick et al., 2009; Camerer et al., 2004; Nagel, 1995). The idea is that players hierarchically form beliefs about their opponents' behavior, up to a certain level of cognitive sophistication (k-level) (see Figure 2). We developed such computational models for hierarchies 1 up to 5 (see Materials and Methods), and first verified that the behavior predicted by different levels of the cognitive hierarchies could be discriminated (see Materials and Methods/Model identifiability and Figure 5). We then fitted those models to our participants' investment data, and ran a Bayesian Model Comparison to identify the hierarchy most likely to generate attacker and defender-like behavior. Our results show that attackers are best described by a model with 4 levels of recursion (model K4, exceedance probability = 67.20%), while defenders are best described by a model with 3 levels of recursion (model K3, exceedance probability = 87.41%) (Figure 5). From these models we estimated, for each subject and each investment in attack and defense, the expected reward, risk prediction, and concomitant reward and risk prediction errors. These reward and risk prediction errors were then related to neural activity, using both whole-brain and ROI-based analyses.

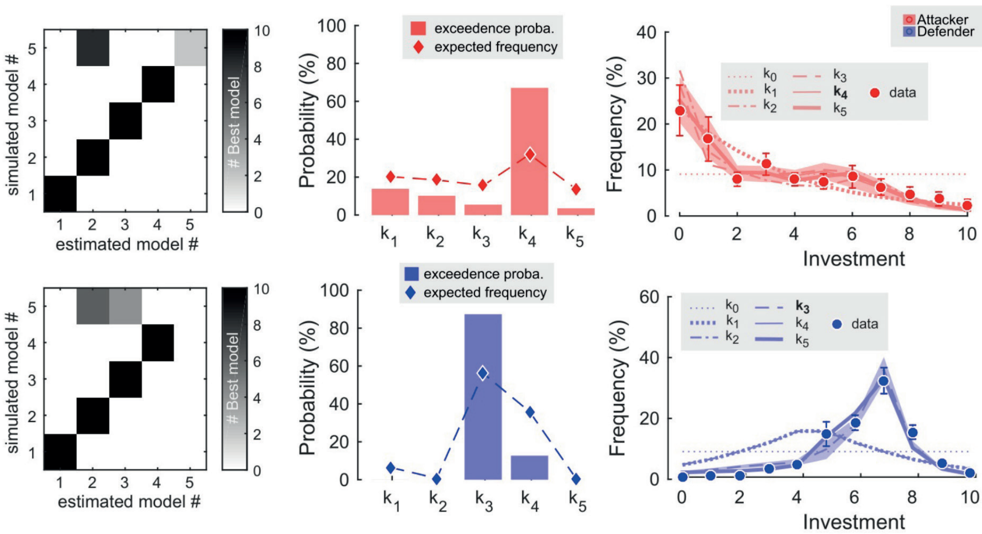

Figure 5. Computational results. (A) Model identifiability, true model used to generate the simulated data (y-axis) and the model estimated as most likely based on our Bayesian Model Comparison (x-axis) for both attacker (top row) and defender (bottom row). (B) Exceedance probability (bars) and estimated model frequencies (diamonds) for both attacker (top row) and defenders (bottom row) of each model fit to participant data. (C) Estimates of each model shown in comparison to true behavioral data for both attacker (top row) and defender (bottom row).

Neural Correlates of Reward Prediction Errors.

Within our VS ROI there was a significant correlation between reward prediction errors and VS neural activity during attack ($t(22) = 2.645$, $p = 0.0148$), but not during defense ($t(22) = -0.330$, $p = 0.745$). Furthermore this correlation between reward prediction errors and VS activity was stronger in attackers than in defenders ($t(22) = 2.189$, $p = 0.0395$, see Figure 6A). Within our amygdala ROI, there was no significant correlation between neural activity and reward prediction errors during either attack ($t(22) = 1.785$, $p = 0.088$), or defense ($t(22) = -1.507$, $p = 0.146$), but there was a significant difference in correlations between the two roles ($t(22) = 2.405$, $p = 0.025$).

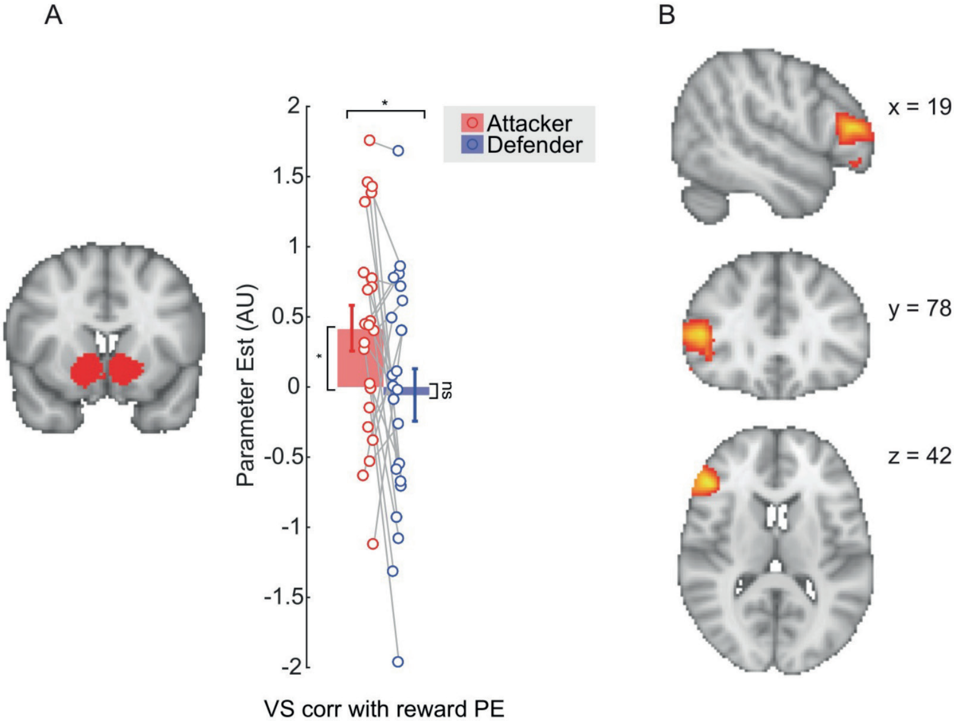

Figure 6. Reward prediction errors differentially relate to attacker and defender neural activity. (A) ROI-analysis reveals prediction errors during attack significantly correlate with ventral striatum activity in attackers but not in defenders. (B) Whole brain analysis reveals that prediction errors during attack significantly correlate with inferior frontal gyrus neural activity. Contrast is FWE-corrected at $p < 0.05$ for the whole brain.

At the whole brain level, we found a cluster in the right IFG that significantly correlated with reward prediction errors during attack (MNI coordinates: x = 48, y = 32, z = 12, Z = 4.55, cluster size = 681, $p = 0.0391$, *FWE*-whole brain, see Figure 6B). We note that this cluster is similar in location to regions found to covary with reaction times (RT), but in the present case the correlation between RT and reward prediction errors was not significant (r = -0.0079, p = 0.654). Because all the contrasts reported were conducted at the feedback time-phase, with the selection time-phase as a co-variate RT was at least partially captured by our GLM. Accordingly, because RT – RPE is non-significant here and RT is captured in the duration of the selection-phase decision-making, we can conclude that RT is not of relevance here.

There were no clusters at the whole brain level that correlated with reward prediction errors during defense, nor were there any clusters that showed a significant difference in correlation between attacker and defender trials.

Neural Correlates of Risk Prediction Errors.

We found that within our VS ROI, there was no significant correlation between neural activity and risk prediction errors during either attack (t(22) = -1.622, p = 0.117), or defense (t(22) = 0.164, p = 0.871), nor was there a significant difference in correlations between the two roles (t(22) = -1.505, p = 0.145). The same was true in our amygdala ROI (attacker: t(22) = -0.588, p = 0.562; defender: t(22) = 0.363, p = 0.720; attacker vs. defender: t(22) = -0.647, p = 0.523) and at a the whole brain level.

Conclusions and Discussion

Competition requires that people expend resources to win from other contestants and to expend resources to prevent losing from other contestants. These two core motives operating during competition – coming out ahead versus not falling behind – were examined here in a simple attacker-defender contest in which opposing individuals simultaneously invested, out of a personal endowment, into exploitative attacks and protective defense. As shown by others already, we find here too that individuals invest less frequently and less intensely in economically “injuring others” than they invest in defending themselves against the threat of being economically injured (De Dreu & Gross, 2019 for a review). Computationally, we found that during attack individuals tend to utilize a higher level of cognitive recursion than during defense. We furthermore found attack behavior relative to defense behavior to be preferentially associated with neural regions associated with theory of mind, and, within the ventral striatum, to be preferentially correlated with reward prediction errors.

What remained poorly understood is why and how people design their strategies of attack and defense. We argued that, in addition to reward maximization, investments in attack and defense may be driven by the desire to out-compete the protagonists as well as by the desire to minimize risk. We approached this issue with a computational framework modeling reward and risk prediction errors based on k-level reasoning in belief formation (Camerer et al., 2004; Nagel, 1995; Zhu et al., 2012). Our results at the neural level revealed no evidence for risk minimization. Instead, and in line with earlier work (e.g., Zhu et al., 2012), we find good evidence that contestants aimed to maximize reward both during attack and defense. At the same time, however, we observed significant differences in the computation of expected reward and in the underlying neural activation during attack versus defense. Specifically, we found reward prediction errors during attack (more than during defense) to robustly correlate with neural activity in the ventral striatum and, using whole-brain analyses, the inferior frontal gyrus.

Our computational modeling demonstrated that investments in attack are best fitted by a model containing four levels of recursion whereas investments in defense are best fitted by a model containing three levels of recursion. This suggests that individuals engage in more sophisticated reasoning about their protagonist's strategy during attack than defense. Indeed, our neuroimaging results revealed significant attack-defense contrasts in neural activation in regions often associated with perspective taking and "Theory of Mind" – the lateral occipital cortex, the inferior frontal gyrus, and the temporoparietal junction (Engelmann, Meyer, Ruff & Fehr, 2019; Prochazkova et al., 2018; Van Overwalle, 2009). These results resonate with earlier work showing that temporarily dysregulating the inferior frontal gyrus through theta burst stimulation affected investment behavior during attack but not defense (De Dreu, Kret, et al., 2016), and that reducing cognitive capacity prior to decision making influenced attackers but not defenders (De Dreu et al., 2019). Combined, these results suggest that individuals engage neural regions for perspective taking and theory of mind during economic contests to out-smart and exploit their protagonist.

Results for neural activity were specific to the feedback phase, when contest outcomes were presented, and not observed during the selection phase when investment decisions were implemented. Possibly, different neurocognitive operations govern implementation and processing of feedback. During implementation, controlled deliberation may be more or less active and this may relate to activity in prefrontal regions involved in executive control. Perhaps the extent to which cognitive control and deliberation during selection is engaged is not conditioned by the specific role decision-makers perform. During feedback, learning and updating operations may be active, and this may relate to neural activation in regions involved in value computation and emotion processing (Behrens, Hunt, & Rushworth, 2009; Yacubian et al., 2006). Indeed, we found neural activity in the ventral striatum to be meaningfully related to reward prediction errors (also see O'Doherty et al., 2004; Stallen et al., 2018; Yacubian et al., 2006; Zhu et al., 2012). In contrast to expectations, however, we did not find differential activity in the amygdala, nor amygdala activity to be related to behavioral indicators processed during feedback. Possibly, contestants process feedback in an emotionally detached and rather cognitive manner aimed at revising and updating their (future) strategy for attack and defense.

Our study design included male participants, and extrapolating conclusions to female participants may be non-trivial. Intuitively competitive success and reward maximization may fit an (evolved) male psychology, whereas risk minimization risk fits an (evolved) female psychology (Croson & Gneezy, 2009; Niederle & Vesterlund, 2011; Spreckelmeyer et al., 2009). At the same time, however, male and female participants tend to perform similarly in the attacker-defender contest studied here (De Dreu & Gross, 2019). Future work is needed to test whether the neurocognitive mechanisms are similar as well, which would further contradict the intuitive hypothesis derived from evolutionary psychology..

Competitions are part and parcel of human life and can be wasteful. In the current contest, subjects destroyed roughly 40% of their wealth in attempts at "injuring others and protecting against being injured" (*viz.* Mill, 1859). Our neurocomputational approach suggested that injuring others is done through rather sophisticated cognitive reasoning, with the key aim to understand the protagonist's strategy selection such that personal rewards can be optimized. When investing in attack more than in defense people engage more sophisticated cognitive recursion. Furthermore, neural structures associated with theory of mind and reward processing are recruited more during attack than defense decisions. Perhaps, mentalizing not only serves empathy and pro-social decision-making, but also the strategic goal of reward maximization through exploitation and subordination.

References

Balodis, I. M., Kober, H., Worhunsky, P. D., Stevens, M. C., Pearlson, G. D., & Potenza, M. N. (2012). Diminished frontostriatal activity during processing of monetary rewards and losses in pathological gambling. *Biological Psychiatry*, *71*(8), 749–757.

Bar-Hillel, M. (2015). Position Effects in Choice From Simultaneous Displays. *Perspectives on Psychological Science*, *10*(4), 419–433. https://doi.org/10.1177/1745691615588092

Baumgartner, T., Heinrichs, M., Vonlanthen, A., Fischbacher, U., & Fehr, E. (2008). Oxytocin Shapes the Neural Circuitry of Trust and Trust Adaptation in Humans. *Neuron*, *58*(4), 639–650.

Behrens, T. E. J., Hunt, L. T., & Rushworth, M. F. S. (2009). The Computation of Social Behavior. *Science*, *324*(5931), 1160–1164. https://doi.org/10.1126/science.1169694

Botvinick, M. M., Niv, Y., & Barto, A. C. (2009). Hierarchically organized behavior and its neural foundations: A reinforcement learning perspective. *Cognition*, *113*(3), 262–280. https://doi.org/10.1016/j.cognition.2008.08.011

Camerer, C., Ho, T. H., & Chong, J. K. (2004). A cognitive hierarchy model of games. *Quarterly Journal of Economics*, *119*(3), 861–898. https://doi.org/10.1162/0033553041502225

Camerer, C., Issacharoff, S. I., Loewenstein, G., Amerer, C. O. C., Ssacharoff, S. A. I., & Oewenstein, G. E. L. (2003). Regulation for conservatives: behavioral economics and the case for "asymmetric paternalism." *University of Pennsylvania Law Review*, (1211), 1211–1254.

Carter, J. R., & Anderton, C. H. (2001). An experimental test of a predator-prey model of appropriation. *Journal of Economic Behavior and Organization*, *45*(1), 83–97.

Charpentier, C. J., Aylward, J., Roiser, J. P., & Robinson, O. J. (2017). Enhanced Risk Aversion, But Not Loss Aversion, in Unmedicated Pathological Anxiety. *Biological Psychiatry*, *81*(12), 1014–1022.

Chen, S., & Bao, F. S. (2015). Linking body size and energetics with predation strategies: A game theoretic modeling framework. *Ecological Modelling*. https://doi.org/10.1016/j.ecolmodel.2015.07.033

Choi, J.-S., & Kim, J. J. (2010). Amygdala regulates risk of predation in rats foraging in a dynamic fear environment. *Proceedings of the National Academy of Sciences*, *107*(50), 21773–21777.

Chowdhury, S. M., Jeon, J. Y., & Ramalingam, A. (2018). Property rights and loss aversion in contests. *Economic Inquiry*, *56*(3), 1492–1511. https://doi.org/10.1111/ecin.12505

Chowdhury, S. M., & Topolyan, I. (2016). The attack-and-defense group contests: best shot versus weakest link. *Economic Inquiry*, *54*(1), 548–557. https://doi.

org/10.1111/ecin.12246

Clark, D. J., & Konrad, K. A. (2007). Asymmetric Conflict. *Journal of Conflict Resolution*, *51*(3), 457–469. https://doi.org/10.1177/0022002707300320

Correa, C. M. C., Noorman, S., Jiang, J., Palminteri, S., Cohen, M. X., Lebreton, M., & van Gaal, S. (2018). How the level of reward awareness changes the computational and electrophysiological signatures of reinforcement learning. *Journal of Neuroscience*, *38*(48), 10338–10348. https://doi.org/10.1523/JNEUROSCI.0457-18.2018

Croson, R., & Gneezy, U. (2009). Gender Differences in Preferences. *Journal of Economic Literature*, *47*(2), 448–474. https://doi.org/10.1257/jel.47.2.448

Daunizeau, J., Adam, V., & Rigoux, L. (2014). VBA: A Probabilistic Treatment of Nonlinear Models for Neurobiological and Behavioural Data. *PLoS Computational Biology*, *10*(1). https://doi.org/10.1371/journal.pcbi.1003441

Daw, N. D. (2011). Trial-by-trial data analysis using computational models. In *Decision Making, Affect, and Learning* (Vol. 6, pp. 3–38). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199600434.003.0001

De Dreu, C. K. W., Giacomantonio, M., Giffin, M. R., & Vecchiato, G. (2019). Psychological constraints on aggressive predation in economic contests. *Journal of Experimental Psychology: General*, *148*(10), 1767–1781. https://doi.org/10.1037/xge0000531

De Dreu, C. K. W., & Gross, J. (2019). Revisiting the form and function of conflict: Neurobiological, psychological, and cultural mechanisms for attack and defense within and between groups. *Behavioral and Brain Sciences*, *42*(2), e116. https://doi.org/10.1017/S0140525X18002170

De Dreu, C. K. W., Gross, J., Méder, Z., Giffin, M., Prochazkova, E., Krikeb, J., & Columbus, S. (2016). In-group defense, out-group aggression, and coordination failures in intergroup conflict. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(38), 10524–10529.

De Dreu, C. K. W., Kret, M. E., & Sligte, I. G. (2016). Modulating prefrontal control in humans reveals distinct pathways to competitive success and collective waste. *Social Cognitive and Affective Neuroscience*, *11*(8), 1236–1244.

De Dreu, C. K. W., Scholte, H. S., van Winden, F. a a M., & Ridderinkhof, K. R. (2015). Oxytocin tempers calculated greed but not impulsive defense in predator-prey contests. *Social Cognitive and Affective Neuroscience*, *10*(5), 721–728.

Dechenaux, E., Kovenock, D., & Sheremeta, R. M. (2015). A survey of experimental research on contests, all-pay auctions and tournaments. *Experimental Economics*, *18*(4), 609–669. https://doi.org/10.1007/s10683-014-9421-0

Delgado, M. R., Schotter, A., Ozbay, E. Y., & Phelps, E. A. (2008). Understanding Overbidding: Using the Neural Circuitry of Reward to Design Economic Auctions. *Science*, *321*(5897), 1849–1852.

Dorris, M. C., & Glimcher, P. W. (2004). Activity in posterior parietal cortex is correlated with the relative subjective desirability of action. *Neuron*, *44*(2), 365–378. https://doi.org/10.1016/j.neuron.2004.09.009

Engelmann, J. B., Schmid, B., De Dreu, C. K. W., Chumbley, J., & Fehr, E. (2019). On the psychology and economics of antisocial personality. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(26), 12781–12786. https://doi.org/10.1073/pnas.1820133116

Flood, M. M. (1972). The Hide and Seek Game of Von Neumann. *Management Science*, *18*(5-part-2), 107–109. https://doi.org/10.1287/mnsc.18.5.107

Goeree, J. K., Holt, C. A., & Palfrey, T. R. (2003). Risk averse behavior in generalized matching pennies games. *Games and Economic Behavior*, *45*(1), 97–113. https://doi.org/10.1016/S0899-8256(03)00052-6

Grossman, H. I., & Kim, M. (1996). Predation and accumulation. *Journal of Economic Growth*, *1*(3), 333–350.

Harsanyi, J. C. (1967). Games with Incomplete Information Played by "Bayesian" Players, I–III Part I. The Basic Model. *Management Science*, *14*(3), 159–182. https://doi.org/10.1287/mnsc.14.3.159

Kahneman, D., & Tversky, A. (1984). Choices. Values. Frames Kahnemann. *American Psychologist*, *39*(4), 341–350. https://doi.org/10.1037/0003-066X.39.4.341

Ku, G., Malhotra, D., & Murnighan, J. K. (2005). Towards a competitive arousal model of decision-making: A study of auction fever in live and Internet auctions. *Organizational Behavior and Human Decision Processes*, *96*(2), 89–103. https://doi.org/10.1016/j.obhdp.2004.10.001

Kuhnen, C. M., & Knutson, B. (2005). The neural basis of financial risk taking. *Neuron*, *47*(5), 763–770. https://doi.org/10.1016/j.neuron.2005.08.008

Lebreton, M., Bavard, S., Daunizeau, J., & Palminteri, S. (2019). Assessing inter-individual differences with task-related functional neuroimaging. *Nature Human Behaviour*, *3*(9), 897–905. https://doi.org/10.1038/s41562-019-0681-8

Loewenstein, G. F., Hsee, C. K., Weber, E. U., & Welch, N. (2001). Risk as Feelings. *Psychological Bulletin*, *127*(2), 267–286. https://doi.org/10.1037/0033-2909.127.2.267

McNamee, D., Rangel, A., & O'Doherty, J. P. (2013). Category-dependent and category-independent goal-value codes in human ventromedial prefrontal cortex. *Nature Neuroscience*, *16*(4), 479–485.

Metereau, E., & Dreher, J. C. (2015). The medial orbitofrontal cortex encodes a general unsigned value signal during anticipation of both appetitive and aversive events. *Cortex*, *63*, 42–54.

Mill, J. S. (1859). *On Liberty*. New York: Walter Scott Publishing.

Mumford, J. A., Poline, J. B., & Poldrack, R. A. (2015). Orthogonalization of regressors in fMRI models. *PLoS ONE*, *10*(4), 1–11. https://doi.org/10.1371/journal.

pone.0126255

Nagel, B. R. (2016). American Economic Association Unraveling in Guessing Games : An Experimental Study Author (s): Rosemarie Nagel Source : The American Economic Review , Vol . 85 , No . 5 (Dec ., 1995), pp . 1313-1326 Published by : American Economic Association Stable , *85*(5), 1313–1326.

Nelson, R. J., & Trainor, B. C. (2007). Neural Mechanisms of Aggression. *Nature Reviews. Neuroscience*, *8*(7), 536–546.

Niederle, M., & Vesterlund, L. (2011). Gender and Competition. *Annual Review of Economics*, *3*(1), 601–630. https://doi.org/10.1146/annurev-economics-111809-125122

Nosenzo, D., Offerman, T., Sefton, M., & van der Veen, A. (2014). Encouraging Compliance: Bonuses Versus Fines in Inspection Games. *Journal of Law, Economics, and Organization*, *30*(3), 623–648. https://doi.org/10.1093/jleo/ewt001

O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., & Dolan, R. J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science (New York, N.Y.)*, *304*(5669), 452–454.

Olsson, A., FeldmanHall, O., Haaker, J., & Hensler, T. (2018). Social regulation of survival circuits through learning. *Current Opinion in Behavioral Sciences*, *24*, 161–167. https://doi.org/10.31234/osf.io/r69e7

Ostrom, E. (1998). A Behavioral Approach to the Rational Choice Theory of Collective Action: Presidential Address, American Political Science Association, 1997. *American Political Science Review*, *92*(01), 1–22. https://doi.org/10.2307/2585925

Palminteri, S., Khamassi, M., Joffily, M., & Coricelli, G. (2015). Contextual modulation of value signals in reward and punishment learning. *Nature Communications*, *6*. https://doi.org/10.1038/ncomms9096

Palminteri, S., Wyart, V., & Koechlin, E. (2017). The Importance of Falsification in Computational Cognitive Modeling. *Trends in Cognitive Sciences*, *21*(6), 425–433. https://doi.org/10.1016/j.tics.2017.03.011

Phelps, E. A., & LeDoux, J. E. (2005). Contributions of the Amygdala to Emotion Processing: From Animal Models to Human Behavior. *Neuron*, *48*(2), 175–187. https://doi.org/10.1016/j.neuron.2005.09.025

Preuschoff, K., & Bossaerts, P. (2007). Adding Prediction Risk to the Theory of Reward Learning. *Annals of the New York Academy of Sciences*, *1104*(1), 135–146. https://doi.org/10.1196/annals.1390.005

Preuschoff, K., Quartz, S. R., & Bossaerts, P. (2008). Human Insula Activation Reflects Risk Prediction Errors As Well As Risk. *Journal of Neuroscience*, *28*(11), 2745–2752.

Prochazkova, E., Prochazkova, L., Giffin, M. R., Scholte, H. S., De Dreu, C. K. W., & Kret, M. E. (2018). Pupil mimicry promotes trust through the theory-of-mind network. *Proceedings of the National Academy of Sciences of the United States of*

America, *115*(31), E7265–E7274. https://doi.org/10.1073/pnas.1803916115

Ribas-Fernandes, J. J. F., Solway, A., Diuk, C., McGuire, J. T., Barto, A. G., Niv, Y., & Botvinick, M. M. (2011). A Neural Signature of Hierarchical Reinforcement Learning. *Neuron*, *71*(2), 370–379. https://doi.org/10.1016/j.neuron.2011.05.042

Rudorf, S., Preuschoff, K., & Weber, B. (2012). Neural Correlates of Anticipation Risk Reflect Risk Preferences. *The Journal of Neuroscience*, *32*(47), 16683–16692.

Smith, S. M., & Nichols, T. E. (2009). Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage*, *44*(1), 83–98.

Spreckelmeyer, K. N., Krach, S., Kohls, G., Rademacher, L., Irmak, A., Konrad, K., … Gründer, G. (2009). Anticipation of monetary and social reward differently activates mesolimbic brain structures in men and women. *Social Cognitive and Affective Neuroscience*, *4*(2), 158–165. https://doi.org/10.1093/scan/nsn051

Stahl, D. O., & Wilson, P. W. (1995). On players' models of other players: Theory and experimental evidence. *Games and Economic Behavior*. https://doi.org/10.1006/game.1995.1031

Stallen, M., Rossi, F., Heijne, A., Smidts, A., De Dreu, C. K. W., & Sanfey, A. G. (2018). Neurobiological Mechanisms of Responding to Injustice. *The Journal of Neuroscience*, 1242–17.

Tobler, P. N., O'Doherty, J. P., Dolan, R. J., & Schultz, W. (2006). Reward Value Coding Distinct From Risk Attitude-Related Uncertainty Coding in Human Reward Systems. *Journal of Neurophysiology*, *97*(2), 1621–1632. https://doi.org/10.1152/jn.00745.2006

Van Overwalle, F. (2009). Social cognition and the brain: A meta-analysis. *Human Brain Mapping*, *30*(3), 829–858. https://doi.org/10.1002/hbm.20547

Winkler, A. M., Ridgway, G. R., Webster, M. A., Smith, S. M., & Nichols, T. E. (2014). Permutation inference for the general linear model. *NeuroImage*, *92*, 381–397.

Wittmann, M. K., Kolling, N., Faber, N. S., Scholl, J., Nelissen, N., & Rushworth, M. F. S. (2016). Self-Other Mergence in the Frontal Cortex during Cooperation and Competition. *Neuron*, *91*(2), 482–493. https://doi.org/10.1016/j.neuron.2016.06.022

Xue, G., Lu, Z., Levin, I. P., Weller, J. A., Li, X., & Bechara, A. (2009). Functional dissociations of risk and reward processing in the medial prefrontal cortex. *Cerebral Cortex*, *19*(5), 1019–1027.

Yacubian, J., Gläscher, J., Schroeder, K., Sommer, T., Dieter, F., & Büchel, C. (2006). Dissociable Systems for Gain- and Loss-Related Value Predictions and Errors of Prediction in the Human Brain. *Journal of Neuroscience*, *26*(37), 9530–9537.

Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., & Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nature Methods*, *8*(8), 665–670. https://doi.org/10.1038/nmeth.1635

Zhu, L., Mathewson, K. E., & Hsu, M. (2012). Dissociable neural representations of reinforcement and belief prediction errors underlie strategic learning. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(5), 1419–1424. https://doi.org/10.1073/pnas.1116783109

3

Chapter 3

Learning to trust through experience and belief

Michael Rojek-Giffin[1,2*], Maël Lebreton[3,4], Andrea Fariña[1,2], Jörg Gross[1,2], and Carsten K.W. De Dreu[1,2,5]

[1] Institute for Psychology, Leiden University, Leiden, the Netherlands
[2] Leiden Institute for Brain and Cognition, Leiden University, Leiden, the Netherlands
[3] Laboratory for Behavioral Neurology and Imaging of Cognition, Department of Basic Neurosciences, Université de Genève, Geneva, Switzerland
[4] Swiss Centre for Affective Sciences, Université de Genève, Geneva, Switzerland
[5] Center for Research in Experimental Economics and Political Decision Making, University of Amsterdam, Amsterdam, the Netherlands

[*] Corresponding author: Leiden Institute of Psychology, Wassenaarseweg 52, 2300 RB, Leiden, The Netherlands: m.r.giffin@fsw.leidenuniv.nl; +31 71 527 6684 or c.k.w.de.dreu@fsw.leidenuniv.nl; +31 71 527 3706

Summary

Trust and reciprocity play core roles in human social interactions. From economic markets to personal relationships, trust and reciprocity allow efficient trade, mutual gain, and cooperation. However, norms of trust and reciprocity differ substantially across individuals and cultures. Here we uncover that variability in reciprocity can be exhaustively captured by three categories: exploiters (individuals who never reciprocate), perfect reciprocators (individuals who always reciprocate), and contingent reciprocators (individuals who reciprocate as a function of how much they are trusted). Due to this variability, trustors are confronted with the challenge to learn who they can trust and who they cannot. Here we investigated this learning process through computational modeling. We show that individuals learn to trust through a combination of reinforcement and belief-based learning. While individuals are able to detect differences between the different reciprocating types through this learning process, they do so sub-optimally. In particular, individuals frequently fail to learn optimal policies towards contingent reciprocators. Furthermore, the degree to which individuals weigh belief over reinforcement is positively correlated to their average payoff, indicating that learning to trust from mentally simulated outcomes outperforms learning from observation alone.

Introduction

Many social transactions afford and require trust and reciprocity. However, norms of trust and reciprocity differ dramatically between cultures (Johnson & Mislin, 2011; Romano, Balliet, Yamagishi, & Liu, 2017) and groups (Heap & Zizzo, 2009; Romano, Balliet, & Wu, 2017), depend on personality traits (Engelmann, Schmid, De Dreu, Chumbley, & Fehr, 2019), and are conditioned by low-level perceptual cues (FeldmanHall et al., 2018; Prochazkova et al., 2018). This variability of reciprocity creates a dilemma for individuals: while trust can lead to mutual gain if investments (e.g. time/energy/money) are reciprocated, trust can also be exploited (Bohnet, Greig, Herrmann, & Zeckhauser, 2008). Therefore, it is imperative that an individual be able to learn who is likely to reciprocate, and who is likely to exploit. Those individuals who consistently err in their predictions lose out whereas those who learn to differentiate the trustworthy from the exploitative can profit from mutually beneficial relationships while avoiding exploitation (Baumard, André, & Sperber, 2013; De Dreu & Gross, 2018). Accordingly, individuals must possess a capacity to learn and predict the reciprocity of others. Here we examine this possibility using computational modeling and behavioral experiments.

Trust has been modeled with economic games such as the trust game (TG) (Berg, Dickhaut, & McCabe, 1995). The TG consists of a sender who decides how much (if any) of a given endowment to transfer (*viz.* entrust) to a responder. The amount transferred to the responder is then increased by some multiplying factor (usually three), after which the responder decides how much (if any) to return (*viz.* reciprocate) back to the sender. Because each unit of the endowment the sender transfers to the responder is tripled, the most collectively profitable outcome involves the sender transferring their entire endowment to the responder. However, in this situation the responder has a strong incentive to exploit the sender's trust and keep the entire sum for themselves. Therefore, the situation that creates the most collective wealth is also the situation that can create the largest inequity and risk of exploitation.

A meta-analysis of TG behavior shows that, on average, senders transfer half of their endowments to responders who, on average, return 40% of the tripled amount back to the sender (Johnson & Mislin, 2011). However, investments as well as amounts returned can vary highly (Balliet & Van Lange, 2013; Johnson & Mislin, 2011), suggesting that different individuals follow different rules when deciding to reciprocate. Indeed, in Experiment 1 and 2 we uncover that nearly all individuals fall into one of three discrete categories: exploiters, perfect reciprocators, and contingent reciprocators. Exploiters are responders who never return as much money to the sender as the sender transferred to them. Perfect reciprocators are responders who always return at least as much money as the sender transferred to them. Contingent reciprocators are responders who return money as a function of how much the sender transferred to them – when the sender transfers a small amount, they return a small amount, and when the sender transfers a large amount, they return a large amount.

These differences in reciprocity naturally lead to questions about trust, namely: how do senders learn who to trust and who to avoid? To answer this question we developed and tested computational learning models. One plausible form this learning process could take is reinforcement learning (RL), which in its simplest form posits that an individual makes a prediction about the value of an action, and updates that prediction based on its outcome (Sutton & Barto, 2018). While RL has been extremely effective in describing human and animal behavior (Behrens, Hunt, & Rushworth, 2009; Erev & Roth, 1998; Palminteri, Wyart, & Koechlin, 2017), and provides a good model for the observed neural processes associated with learning (Behrens, Hunt, Woolrich, & Rushworth, 2008; Levy & Glimcher, 2012; Rutledge, Dean, Caplin, & Glimcher, 2010), in many settings RL makes untenable assumptions about how individuals actually reason and adapt to the environment. One of these assumptions is that individuals only update actions they select, and ignore all other strategies available to them.

A more psychologically plausible account makes the claim that individuals in fact simulate the outcome of the other strategies available to them, and update all strategies based on these simulated outcomes. This so-called belief-based learning (BB) is mathematically equivalent to Bayesian updating under specific assumptions (Fudenberg & Levine, 1998; Zhu, Mathewson, & Hsu, 2012), and generalizes RL to include beliefs and counterfactual action simulation. However, BB posits that all strategies are updated with equal weight, whether they were selected or not. This assertion that mentally simulated outcomes are treated identically to experienced outcomes is as untenable as the assertion from RL that mentally simulated outcomes are ignored.

A third alternative offers a unification of these two approaches through a hybrid model which allows for each of these separate accounts as special cases (Camerer & Ho, 1999). This model, Experience Weighted Attraction (EWA), has been shown to account for behavior in a variety of economic games better than RL or BB models (Camerer & Ho, 1999; Camerer, Ho, & Chong, 2002; Ho, Camerer, & Chong, 2007; Zhu et al., 2012). Furthermore, the parameters of the EWA model have specific psychological interpretations, allowing for insights into the inner workings of individuals' decision-making process that exceed those allowed by simpler models.

Here we apply these different models to understand how people learn to trust. We show that when individuals play the TG as senders against the three different responder categories (exploiter, perfect reciprocator, contingent reciprocator), the EWA model captures behavior better than RL and BB alternatives. This means that when faced with the dilemma of learning who to trust and who to avoid, individuals combine their own experiences with their subjective beliefs about others. Interestingly, we further showed that reliance on belief was positively correlated with average payoff, indicating that mental simulation during the process of learning to trust, although only partially employed, affords a direct benefit to the individual.

Experiments 1 and 2

Materials and Methods

Ethics and subject recruitment

Experiments received ethics approval from Leiden University (Exp. 1: CEP19-0131/40; Exp. 2: CEP19-0108/7). Participants (Exp. 1: N = 272; Exp. 2: N = 106) were recruited from MTurk (Exp. 1) or the subject pool at Leiden University (Exp. 2), provided digital or written informed consent, and were debriefed and paid for participation. All experiments were incentivized and did not involve deception. Individual anonymity was guaranteed throughout and earnings were paid in private.

The Trust Game

To investigate trust, we used the dyadic trust game (TG) (Berg et al., 1995; Johnson & Mislin, 2011) that is played between a sender and a responder. Senders began each trial with an endowment of e = 20 monetary units (MU) (1MU = €0.025). The sender could transfer any amount t (in steps of 1MU) to the responder ($t \in [0,20]$). Any amount transferred to the responder was tripled ($3t$). The responder was then allowed to return any amount r between 0 and $3t$ (in steps of 1MU) back to the sender ($r \in [0,3t]$). At the end of each trial, each player received their MU as their payoff (sender payoff = $e - t + r$, responder payoff = $3t - r$).

Because any amount transferred by the sender is at risk of being kept by the responder, the amount transferred is a measure of trust. On the other hand, because the responder is under no obligation to return anything to the sender, any amount they do return is a measure of reciprocity (or the return of trust). All players were paid out based on two randomly selected trials (resulting in a maximum bonus payout of €3.00).

Detecting different types of reciprocators

In order to categorize subjects' reciprocation behavior, we collected data from participants playing the TG as responders using the strategy method. Specifically, participants were asked what they would return for all possible transfers from the sender. Participants were told that their data would be used as feedback for senders in a future experiment, and that, in addition to the payment they received after the experiment, they would receive additional bonuses every time their responses were used in subsequent experiments.

The canonical view of rationality from economics would predict that responders never return anything to senders (Henrich et al., 2005). However, this view has been consistently refuted by studies showing humans to be cooperative (De Dreu & Gross, 2018), fairness seeking (Fehr & Schmidt, 1999), and innately prosocial (Rand, Greene, & Nowak, 2012). Based on this view, we would predict that responders generally act "fairly" towards senders. However, the concept of fairness differs across individuals and cultures (Bohnet et al., 2008; Henrich et al., 2005). One intuitive conceptualization of fairness posits that a responder always returns back to the sender roughly half of what they receive. However an alternative view posits that responders are attempting to approximate a fair division of the total amount of money available on any given trial. Based on this view, we would expect that when a sender transfers a low amount and keeps most of the endowment for themselves, the responder reacts by returning a low amount; conversely we would expect that when a sender transfers a large amount, the responder reacts by returning a large amount, hence approximating an even split of the total sum available. Importantly, this conceptualization predicts responders to reciprocate in an ascending fashion. This reasoning led us to conduct two complementary analyses.

In the first analysis, we dividing our responders into those who never returned as much as the sender transferred ($\forall r < t$), those who always returned as much or more than the responder transferred ($\forall r \geq t$), and those who returned low transfers with less than the transferred amount and high transfers with more than the transferred amount,

$$\begin{cases} \forall r < t, \ if \ t = 1 \\ \forall r \geq t, \ if \ t = e \end{cases}.$$

We confirmed this categorization using a data-driven approach. In this analysis, we fit logistic functions to each subject's individual response data, with each logistic function having 4 free parameters: slope, inflection point, lower plateau, and upper plateau. We then conducted a weighted principal component analysis on the fitted parameters. Within each component, we ranked each subject based on their PC-score, indicating how much variance was explained by their particular combination of parameters relative to other subjects within the given component. We then divided subjects based on this PC-score into three equal groups within the component. Finally, we plotted the average return for each transfer within each of these groups. The patterns that emerged clearly qualitatively mimicked those of the above described return-rate division.

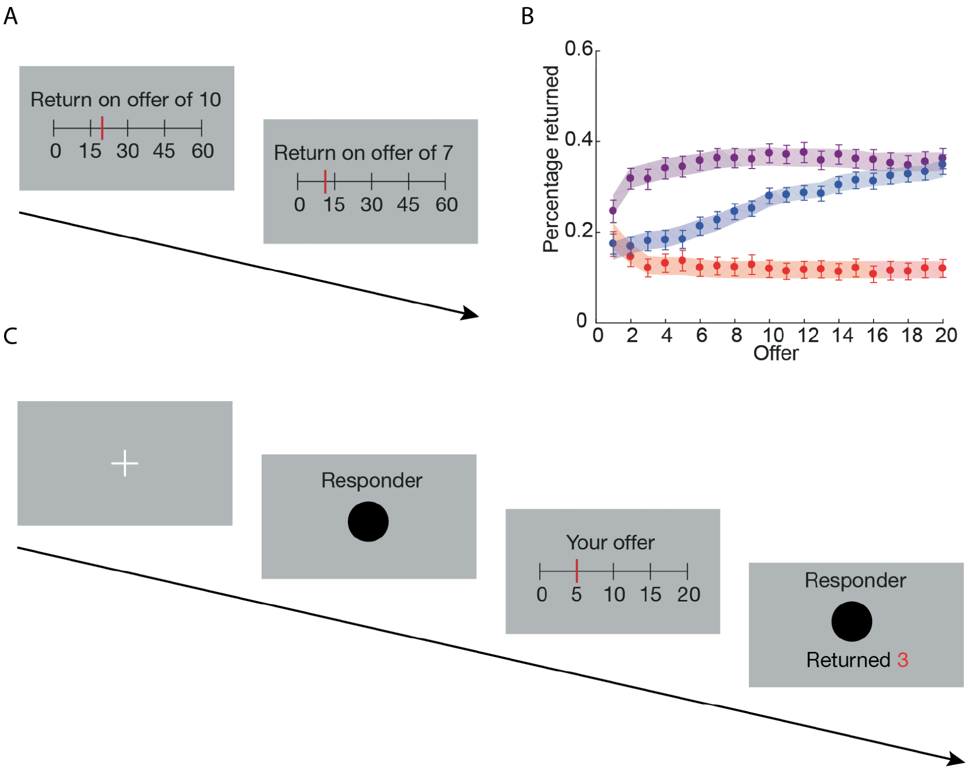

Figure 1. Schematic of game and timeline. (A) Trial timeline for the behavioral task in Experiments 1 and 2. Participants in the role of responder were asked to return an amount for every possible amount transferred to them by the sender. This resulted in (B) a division of responders into three categories: exploiters (red line), perfect reciprocators (purple line), and contingent reciprocators (blue line). This dots with error bars are the data, and the shaded area are model fits (see Experiments 1 and 2 – Material and Methods/ Detecting different trustworthy type). These responder categories were used to provide feedback for (C) Experiments 3 and 4, during which senders made transfers to each responder type, each of which was identified with a neutral shape.

Results

Trust responders

We found that the majority of participants (Exp. 1: N = 219, 80.51%; Exp. 2: N = 94, 88.68%) could be categorized as one of three distinct reciprocity types: individuals who never fully reciprocated, so-called exploiters (Exp. 1: N = 78, 28.68%; Exp. 2: N = 18, 16.98%); individuals who always reciprocated, so-called perfect reciprocators (Exp. 1: N = 94, 34.56%; Exp. 2: N = 43, 40.57%), and individuals who reciprocated as a function of how much they were transferred, so-called contingent reciprocators (Exp. 1: N = 47, 17.28%; Exp. 2: N = 34, 33.02%) (Fig. 2).

We validated our pre-defined categorization with a data-driven approach using principle component analysis (PCA; see *Material and Methods/ Detecting different* *trustworthy type*). This analysis corroborated our theory-driven approach, with subjects exhibiting return patterns consistent with either exploiter, perfect reciprocator, or contingent reciprocator (see Fig. 2). In short, the vast majority (>80%) of the sample fell into one of these three discrete categories of reciprocity.

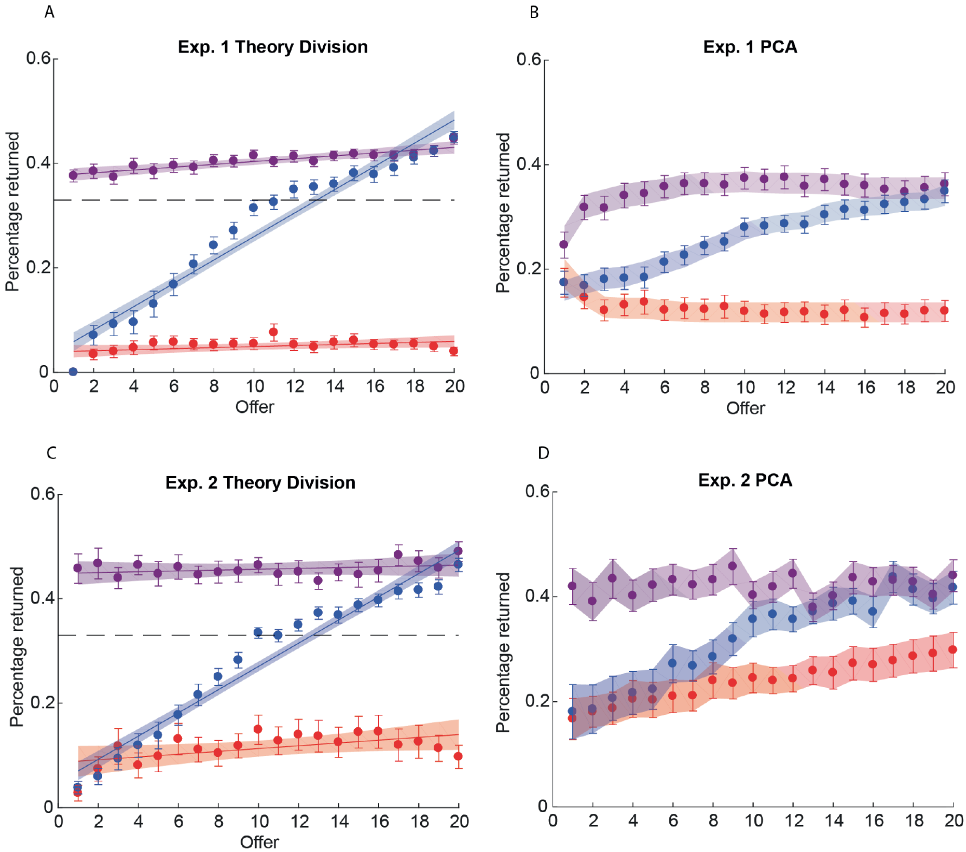

Figure 2. Different trustworthy types in both theory and data driven analyses. (A, C) Theory driven division of responders around "fair" return (dotted line), which would return sender to their original endowment. The three lines represent responders who either always return as much or more than what was transferred to them (purple), responders who always return less than what was transferred to them (red), and responders who return low transfers with less but high transfers with more than what was transferred to them (blue). (B, D) Data-driven division of responders through principal component analysis. We fit logistic functions to response data, subjected the parameters of these logistic fits to PCA, and then split up the results into three discrete groups within PC-space. The resulting division clearly qualitatively matched the results obtained from our theory-driven approach.

Summary and Discussion for Experiment 1 and 2

The results of Experiments 1 and 2 converge on the notion that reciprocity behavior can be exhaustively captured within three simple categories: exploiters, perfect reciprocators, and contingent reciprocators. This tripartite division may at first seem counterintuitive, since it is tempting to think of behavior as either fair or not and thus to observe individuals who are either reciprocal or not. This reasoning should lead to a binary division of responders into exploiters (individuals who never adhere to norms of fairness) and reciprocators (individuals who always adhere to norms of fairness). However, what this binary division ignores is the multifarious nature of the concept of fairness, even within a given culture or group (Henrich et al., 2005). We attempted to capture this multifarious nature by considering the possibility that individuals may reciprocate based on different rules. Accordingly, in both Experiments 1 and 2 we found that, in addition to acting exploitatively, responders acted either as "perfect reciprocators" (returns were consistent across transferred amount), or "contingent reciprocators" (returns were contingent on transferred amount). While still simplistic, these three categories captured the vast majority of our sample (>80%). This was given credence by an initial theory driven approach, and a subsequent principal component analysis. Furthermore, this was replicated across both online and laboratory studies, suggesting that these different types are composite elements of the population. If these different reciprocating types are indeed as common as these experiments imply, then we should expect to see them readily learned by naïve individuals in the population. This reasoning led us to conduct Experiments 3 and 4.

Experiments 3 and 4

Materials and Methods

Ethics and subject recruitment

Experiments received ethics approval from Leiden University (Exp. 3: CEP19-0108/7; Exp. 4: CEP19-0131/40). Participants (Exp. 3: $N = 98$; Exp. 4: $N = 106$) were recruited from MTurk (Exp. 3) or the subject pool at Leiden University (Exp. 4), provided digital or written informed consent), and were debriefed and paid for participation. Exp. 2 and Exp. 4 consisted of the same participants and were conducted in the same experimental session, with participants always first making decision in the TG as responders and subsequently making decisions as senders. All experiments were incentivized and did not involve deception. Individual anonymity was guaranteed throughout and earnings were paid in private.

Learning to trust

Participants played multiple rounds of the TG as senders to the three different responder categories found in Exp. 1 and 2 (see *Experiments 1 and 2 – Results/Trust Responders*). To provide feedback to senders, we used the results of the PCA analysis from Exp. 1 (see *Results/Trust Responders*). Specifically, for each amount the sender transferred, they were returned an amount within the interval of the return percentage specific to that particular responder, displayed as the shaded areas on Fig. 1B. For example, if a sender transferred 10MU to a contingent reciprocator (blue line), he or she would receive back a random amount between 25% and 29% of the tripled amount, which resulted in a return between 7MU and 9MU. This method allowed us to provide feedback to senders that maximized noise (so as not to have feedback entirely deterministic) while still maintaining the same division of responders into exploiters, perfect reciprocators, and contingent reciprocators. Subjects played 4 blocks of 36 trails (Exp. 3) or 2 blocks of 72 trials (Exp. 4) against each responder category. Responders were identified by a neutral shape (e.g., a square, a circle, or a triangle), with shapes randomized for each participant. To ensure the independence of learning within each block, every block consisted of completely novel shapes.

Computational modeling

Reinforcement and Belief-Based Learning. We utilized computational modeling in order to gain insights into the mechanisms behind sender decision-making. One plausible form decision-making could take is reinforcement learning (RL), which in its simplest form posits that individuals make predictions about the value of an action, and update those predictions based on their outcomes (Sutton & Barto, 2018). More specifically, every given action, $x(t)$, has an associated expected value $V(x(t))$. On every trial t, a prediction error is calculated, which is the actual outcome subtracted from the expected outcome:

$$PE = EV\big(x(t)\big) - outcome$$

This prediction error is then used to update the value of that action using an additional weighting parameter α acting as the learning rate:

$$\big(x(t+1)\big) = EV\big(x(t)\big) + \alpha \times PE$$

While this simple formulation has been extremely effective in describing decision-making, in some situations it oversimplifies the process and relies on untenable assumptions. One of these assumptions is that individuals only update their selected action, and leave all unselected options unchanged.

An alternative account makes the claim that individuals in fact simulate the outcome of the other strategies available to them, and update all strategies based on these simulated outcomes. This so-called belief-based learning (BB) expands RL to include beliefs about unselected actions, such that every possible action is updated on each trial. However, BB posits that all strategies are updated with equal weight, whether they were selected or not, which is an assumption just as untenable as those made by RL. A third alternative offers a unification of these two models into a single framework.

Experience Weighted Attraction. The experience weighted attraction (EWA) framework combines reinforcement learning (RL) and belief-based learning (BB) into a single hybrid model (Camerer & Ho, 1999). In its original form, the EWA model contains two key variables which are updated on each trial: $N(t)$, and $A_i^j(t)$. The variable $N(t)$ is the number 'observational-equivalents', which is to say the number of times a subject has experienced, or believes they have experienced, an interaction with a given opponent. The variable $A_i^j(t)$ is a vector of "attractions", which is the value ascribed to every possible option. On each trial, $N(t)$ is updated according to the following rule:

$$N(t) = \rho \times N(t-1) + 1$$

Where p is a free parameter which controls the depreciation of $N(t)$ over trials. The variable $A_i^j(t)$ is updated according to the following rule:

$$A_i^j(t) = \begin{cases} \dfrac{\phi \times N(t-1) \times A_i^j(t-1) + reward}{N(t)}, & if A_i^j(t) = x(t) \\ \dfrac{\phi \times N(t-1) \times A_i^j(t-1) + (\delta \times reward)}{N(t)}, & if A_i^j(t) \neq x(t) \end{cases}$$

Where $x(t)$ is the option selected by the participant, i.e. the amount of money transferred to the responder. The free parameter ϕ controls the depreciation of past attractions for options that the participant selected, and the free parameter δ controls the depreciation of past attractions for options that the subject did not select.

In our framework, we assumed that senders formed a priori estimates of each option's value by estimating the function governing each responder category's return-rate for each option. These so-called response functions took a linear form and consisted of two additional free parameters, a prior slope and prior intercept, which provided the initial attractions, $A_i^j(0)$. Based on previous literature, we set $N(0)$ to 1 (Ho et al., 2007). As is common in the computational modeling literature (Daw, 2011), we modeled actual choice selection with a softmax function, which included an additional free parameter β, which is the inverse temperature parameter controlling the participant's exploration/exploitation trade-off. Lower values of β indicate more exploratory (stochastic) behavior, while higher values indicate more exploitative (deterministic) behavior.

The primary benefit of the EWA model is that it allows for both RL and BB learning simultaneously. These two forms of learning are controlled by δ, with higher values of δ indicating a higher reliance on BB learning and as such a higher reliance on simulating the outcomes of unselected options (so-called forgone outcomes). To simulate the outcome of an unselected option, we reasoned that participants could (I) use their prior estimates of the responders' response function, (II) extrapolate from the observed percentage returned by the responder, or (III) extrapolate from the observed raw amount returned by the responder. We therefore designed different versions of the EWA model in which unselected options were updated based one each of these possibilities. Furthermore, participants could be weighing unselected options asymmetrically. We therefore devised models in which options above and below the chosen option had their values updated by two different δ parameters: δ^- and δ^+ . We also tested for the possibility that participants were weighing forgone outcomes from each different responder category with a different weight, resulting in three δ parameters. Finally, as stated above, the EWA model allows for both RL and BB learning as special cases. When $\delta = 0$ and $\rho = 0$, the model is equivalent to RL (with ϕ being equivalent to the learning rate α), and when $\delta = 1$ and $\rho = \phi$, the model is equivalent to complete BB learning (forgone and experienced outcomes are weighted equally) (Camerer & Ho, 1999). For this reason, we added three BB models (one for each foregone outcome method), and one RL model. Therefore we tested a total of 13 models: 3 (number of $\delta's$) $\times$ 3 (forgone outcome update method) + 3 (BB with different foregone outcome methods) + 1 (RL).

Model Fitting Procedure.

We optimized each models' free parameters by minimizing the negative log likelihood (LLmax) of the participant's observed choices under the model using Matlab's fmincon function, initialized at multiple starting points of the parameter space. Negative log likelihoods were used to compute, at the individual level and for each model, the Akaike Information Criterion (AIC), which was used to quantify model evidence $(me = -AIC/2)$ (Correa et al., 2018). We then employed a random effects Bayesian model comparison using the Variational Bayesian Analysis (VBA) toolbox (Daunizeau, Adam, & Rigoux, 2014) to estimate the exceedance probability (denoted XP) for each model. Exceedance probability quantifies the evidence that the model is more likely than all the other models tested. An exceedance probability greater than 95% for one model within a test-set is therefore typically considered as significant evidence in favor of this model being the most likely. In our case, even after accounting for the additional free parameters, and with a total of 13 models tested against one-another (risking probability dilution), the EWA model which simulated foregone outcomes by extrapolating from the observed percentage returned, and with asymmetrical deltas, was significantly better than all other models tested, including both the RL and the BB alternatives (all XP's > 95%, see Fig. 4).

Parameter recovery and model identifiability.

As a further validation, we ran simulations which mimicked the experimental parameters of both Exp. 3 and 4. In this procedure, we generated choice behavior for 98 and 106 simulated subjects for each of our 13 models. Parameters were randomly sampled from probability distributions which approximated the distribution of parameters estimated from fitting the EWA model to the choices of senders in both Experiments 3 and 4. The experimental constraints were identical to those of each experiment, that is, 98 and 106 simulated subjects playing 4 blocks of 12 trials and 2 blocks of 24 trials, respectively, for each of our 13 models. We then estimated the parameters of the model that was used to simulate the behavior based solely on the simulated behavior. Finally, we correlated the estimated model parameters with the simulated model parameters. For each model, the correlation between the true and estimated parameters was highly significant (all p's < 0.001; Fig. 3CD), indicating that each model's parameters could be reliably estimated. Finally, Variational Bayesian Analysis (Daunizeau et al., 2014) was used to confirm whether the model used to simulate the data was also selected as the most likely model based on our goodness of fit measure (AIC). In each case the model used to generate the data was selected as the most likely (XP's > 95%; Fig. 3AB).

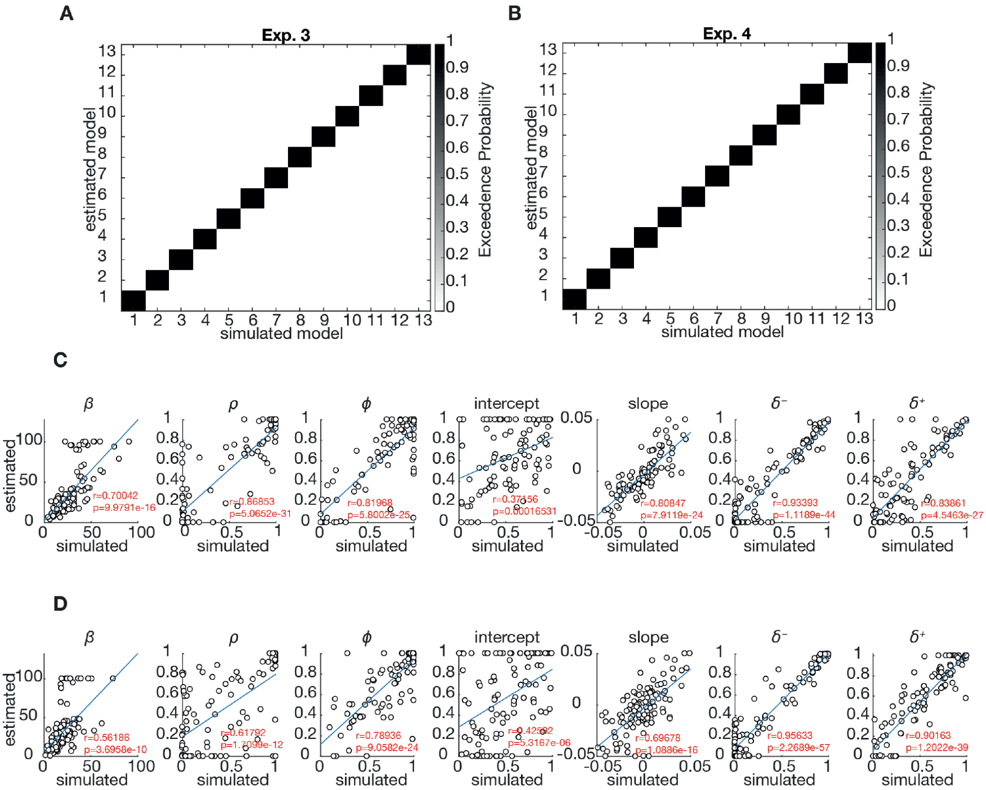

Figure 3. Model identifiability and parameter recovery. We ran simulations which mimicked the experimental parameters of both Exp. 3 (A, C) and Exp. 4 (B, D), that is, 98 and 106 subjects playing 4 blocks of 12 trials and 2 blocks of 24 trials, respectively, for each of our 13 models. Parameters were randomly sampled from probability distributions which approximated the distribution of parameters estimated from fitting the EWA model to the choices of senders in both Experiments 3 and 4. We then estimated the model (A, B) and the parameters (B, C) of the model used to simulate the behavior based solely on the simulated behavior. Panels (A, B) show the exceedance probability of all models fit to each other, with the dark diagonals indicating that in each case the model used to simulate the data was the model identified as most likely. Panels (A, B) show the correlation between the simulated and estimated parameters for each parameter from our winning model (model 6, EWA with different δ parameters for options above and below selected option.).

Results

Trust senders.

Model-free results. Over trials, subjects did adapt their transfers to the different responder types over trials (note: exploiter dummy coded as baseline; Exp. 3: perfect reciprocator × trial: $b \pm$ se = 0.585 ± 0.036, $p < 0.001$, contingent reciprocator × trial: $b \pm$ se = 0.363 ± 0.036, $p < 0.001$; Exp. 4: perfect reciprocator × trial: $b \pm$ se = 0.260 ± 0.015, $p < 0.001$, contingent reciprocator × trial: $b \pm$ se = 0.167 ± 0.015, $p < 0.001$) (see Table 1). This demonstrates that learning was indeed taking place, however this learning occurred differentially depending on the responder category a sender faced (Fig. 5).

Table 1: Multi-level Regression on sender transfers over trials.

	Experiment 3	Experiment 4
	B (se)	B (se)
Trial	-0.409 (0.025)***	-0.151 (0.011)***
Perfect Reciprocator	4.260 (0.263)***	4.708 (0.221)***
Contingent Reciprocator	2.652 (0.263)***	2.040 (0.221)***
Perfect Reciprocator × Trail	.585 (0.036)***	0.260 (0.015)***
Contingent Reciprocator × Trail	0.363 (0.036)***	0.167 (0.015)***
Observations	98	106

Note. Exploiter dummy coded as baseline. Standard Errors are clustered within subjects (N = 98 for Exp. 3, and N = 106 for Exp. 4). *** $p < 0.0001$, ** $p < 0.001$, * $p < 0.05$, # $p < 0.10$ (two-tailed tests).

Interestingly, subjects on average behaved sub-optimally against all responder categories. An optimal transfer to both the perfect and contingent reciprocators was 20MU (which would yield a return rate between 32% and 38%), and an optimal transfer for the exploiters was 0MU, since no transfer would return the sender to their original endowment. However, transfers on the final trial were significantly different from the optimum in all cases (Exp. 3: exploiters: $t(97) = 9.272$, $p < 0.001$; perfect reciprocators: $t(97) = -13.265$, $p < 0.001$; contingent reciprocators: $t(97) = -19.455$, $p < 0.001$; Exp. 4: exploiters: $t(105) = 7.899$, $p < 0.001$; perfect reciprocators: $t(105) = -10.814$, $p < 0.001$; contingent reciprocators: $t(105) = -16.497$, $p < 0.001$). Repeated-measures ANOVAs (with Huyn-Feldt corrections for violations of sphericity) with participant distance from optimum on the final trial $(abs(optimum - final\,offer))$ as the dependent variable and responder category as the independent variable revealed that the degree of suboptimality differed significantly between the different responder types (Exp. 3: $F(2, 194) = 55.599$, $p < 0.001$, partial $\eta^2 = 0.223$; Exp. 4: $F(2, 210) = 50.844$, $p < 0.001$, partial $\eta^2 = 0.215$). Furthermore, participants were farthest from optimal when interacting with contingent reciprocators (Exp. 3: exploiters vs. perfect reciprocators: $t(97) = -5.605$, $p < 0.001$; exploiters vs. contingent reciprocators:

Chapter 3

$t(97) = -9.238$, $p < 0.001$; perfect reciprocators vs. contingent reciprocators: $t(97) = -5.937$, $p < 0.001$; Exp. 4: exploiters vs. perfect reciprocators: $t(105) = -3.962$, $p < 0.001$; exploiters vs. contingent reciprocators: $t(105) = -8.697$, $p < 0.001$; perfect reciprocators vs. contingent reciprocators: $t(105) = -7.260$, $p < 0.001$). In other words, participants received less money than they could have if they had adapted more strongly to the responder categories, particularly to the contingent reciprocators.

Model-based results.

We next assessed the mechanisms responsible for the learning process participant's utilized when interacting with these different responder categories. To this end, we tested several computational models that were psychologically plausible and shown to effectively explain empirical data in economic games (Camerer & Ho, 1999; Erev & Roth, 1998).

In total, we constructed 13 candidate models (see *Materials and Methods/ Computational modeling/Experience Weighted Attraction*) which could account for learning our TG set-up. Bayesian model comparison revealed the winning model to be the one that extrapolated the experienced percentage returned to the rest of the option space (Exp. 3: XP = 98.76%; Exp. 4: XP = 100%; see Fig. 4). We examined this model's fit to the empirical data for both Experiments 3 and 4, and in both cases this version of the EWA model did captured a significant amount of sender behavioral variance, (Exp. 3: $R^2 = 0.339$; Exp. 4: $R^2 = 0.351$; see Fig. 5). This reveals that subjects were more cognizant of the percentage returned than the total amount returned, suggesting they were more concerned with the amount they received relative to the responder than the amount they received in total.

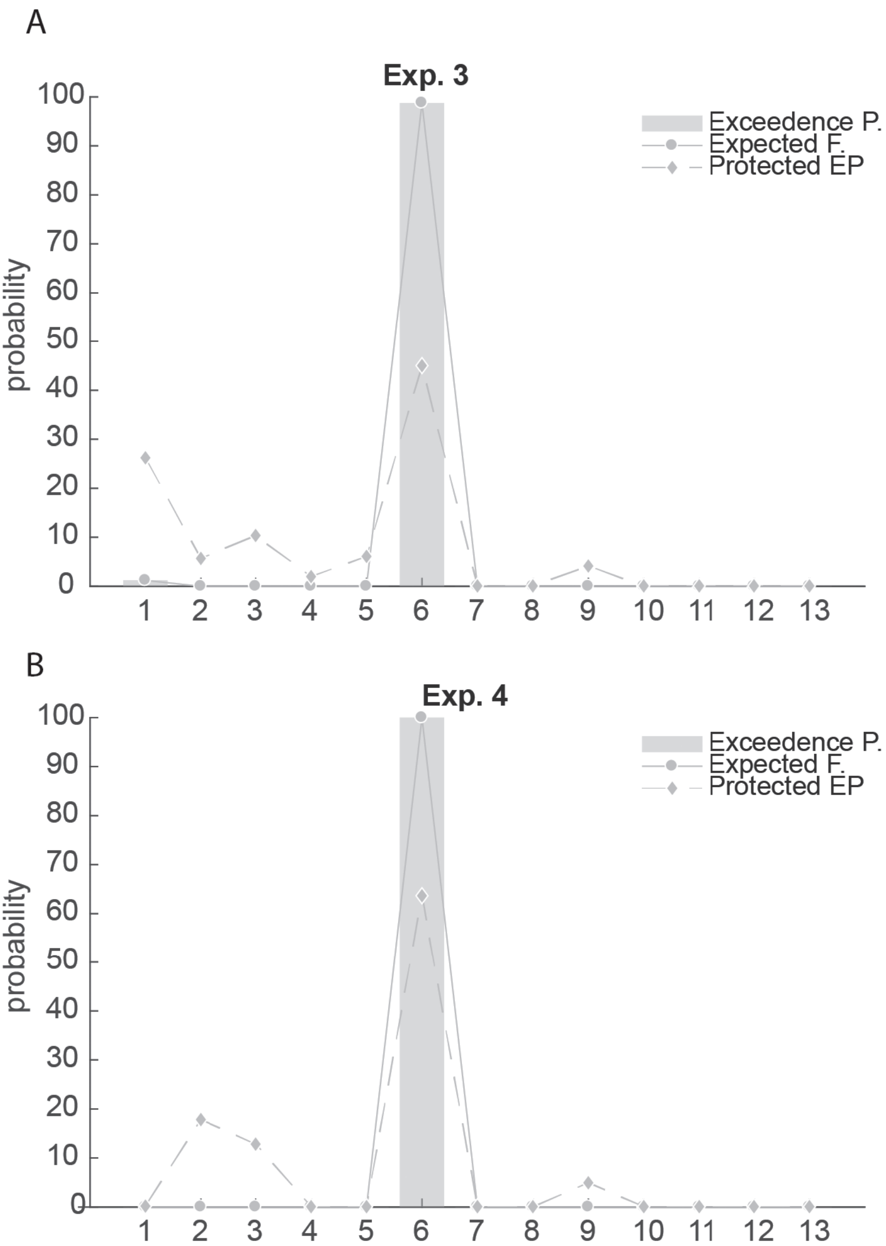

Figure 4. Bar-plots of model comparison. We used Variational Bayesian Analysis (Daunizeau et al., 2014) in order to compare our candidate models. Shown are results from this model selection procedure for both Exp. 3 (A) and Exp. 4 (B). For both datasets, the winning model was model 6: experience weighted attraction, estimating payoffs from unselected options by extrapolating from the percentage returned of the selected option, and with different δ parameters for unselected options above and below the option selected. For descriptions of all other models, see Materials and Methods/Computational Modeling/Experience Weighted Attraction.

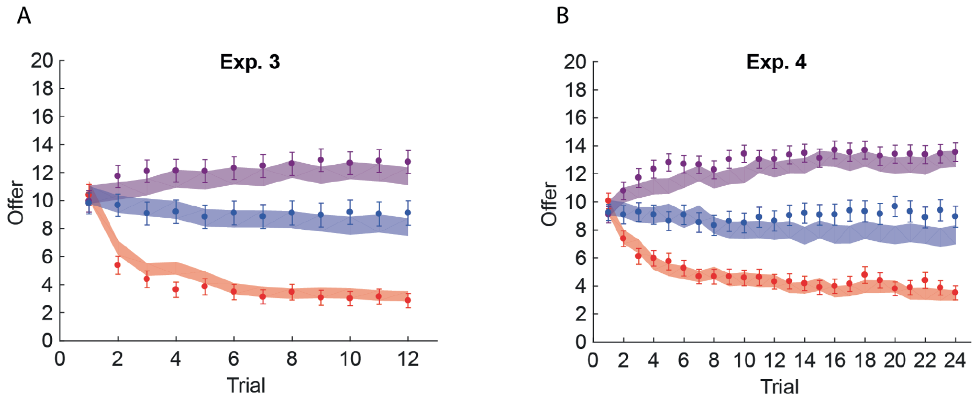

Figure 5. Learning curves with model fits to data. Across trials, participants adjusted to the different reciprocity types differentially (exploiter = red, perfect reciprocator = purple, contingent reciprocator = blue), a process that was well captured by our modeling approach. Dots with error bars represent mean ± standard error, shaded areas represent model predictions for Experiment 3 (A) and Experiment 4 (B).

Furthermore, the winning model updated forgone outcomes differentially, operationalized by two δ parameters which differentially accounted for updating unselected options above and below the chosen option. Options below the selected option were updated with a significantly lower weight than those above the selected option for Exp. 4 ($t(105) = -4.315$, $p < 0.001$), however this effect was absent in Exp. 3 ($t(97) = 1.213$, $p = 0.229$; Fig. 6AB), possibly due to the lower number of trials in this experiment relative to Exp. 4. Because δ denotes reliance on BB relative to RL learning, one interpretation of δ is that participants were more apt to simulate the outcomes of options that could yield higher payoffs than those they actually received. This can be seen as an optimism bias, with subjects showing a higher propensity for imagining brighter relative to darker alternative outcomes.

This reliance on BB learning also had tangible effects on sender outcome. The higher a sender's δ (both positive and negative), and thus the more they relied on BB learning, the higher they earned overall (Exp. 3: δ^-: $R = 0.689$, $p < 0.001$, δ^+: $R = 0.502$, $p < 0.001$; Exp. 4: δ^-: $R = 0.657$, $p < 0.001$, δ^+: $R = 0.284$, $p = 0.005$, Fig. 6C-F). In other words, the more an individual relied on learning from fictitious play, the more money they received in this task. This means that simulated learning was actually beneficial for learning to trust, despite the fact that it was only partially employed by participants.

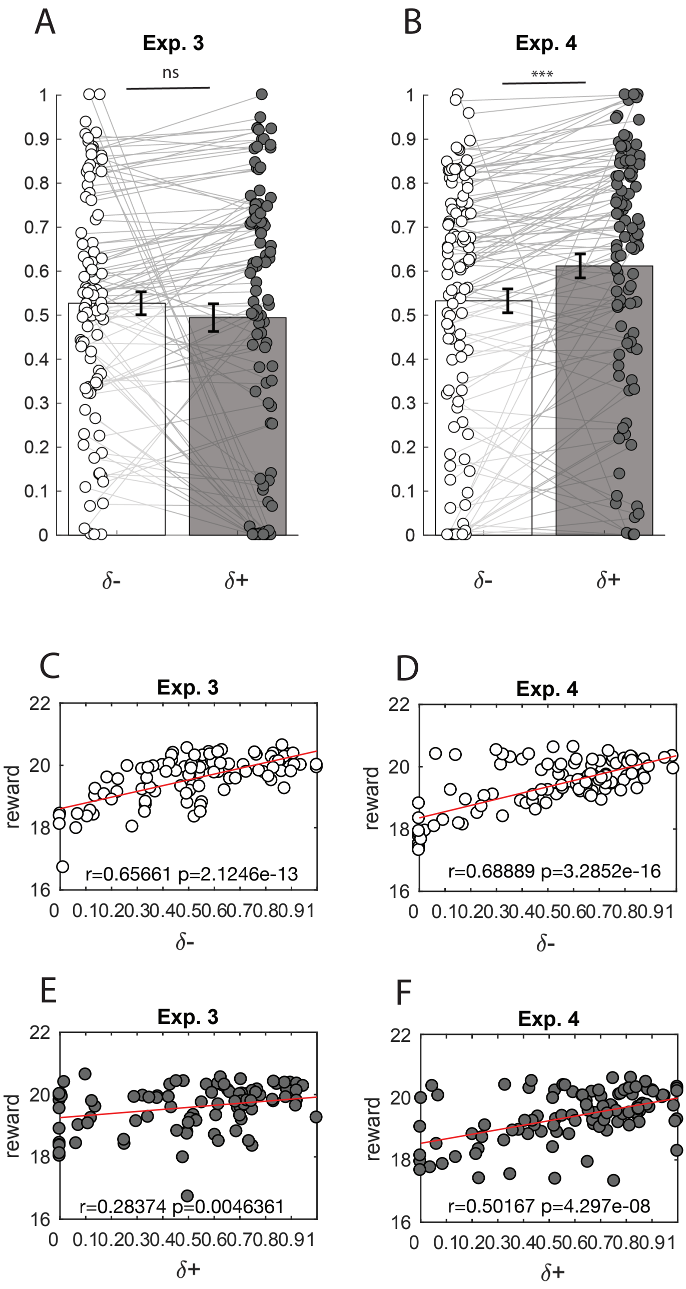

Figure 6. Comparing δ above and below selected option and correlation between δ and reward . For Exp. 4 (B), but not Exp. 3 (A), the parameter δ, which controls the tradeoff between reinforcement and belief-based learning strategies, differed significantly, meaning unselected options above selected options had their values adjusted with a higher weight than options below selected options. Additionally, the parameter δ robustly correlated with average reward across subjects in both Exp. 3 (C, E) and Exp. 4 (D, F). This was true for δ's adjusting weights below (C, D) and above (E, F) the selected option in both data sets. Each dot represents one subject. Bars and error-bars represent mean and SEM, respectively.

Discussion

Trust is at the core of many human interactions within economic, political, and social spheres. Engaging in a trusting relationship can be profitable to all parties involved, however trusting behavior is always at risk of being exploited. Thus, it is crucial to understand the different trustworthy characteristics that exist in the general population, and how these different characteristics are learned. By collecting data from both responders and senders in the classic two-step trust game (Berg et al., 1995), we show that three simple rules of trustworthiness can exhaustively capture responder behavior, and that these different rules are (suboptimally) learned by naïve senders through a combination of reinforcement and belief-based learning.

When considering trustworthy (*viz.* responder) behavior, it is tempting to fall into a trap of binary thinking: that an individual is either selfish (exploiter) or not (perfect reciprocator). Indeed while economic theory predicts that individuals should always exploit in a one-shot interaction such as that used in the current setup, humans in reality are cooperative (De Dreu & Gross, 2018), fairness seeking (Fehr & Schmidt, 1999), and prosocial (Rand et al., 2012). Therefore we should expect that at least some individuals behave "fairly", despite the fact that interactions are one-shot and anonymous. However, "fairness" itself differs across individuals and cultures (Bohnet et al., 2008; Henrich et al., 2005), and can be manifested in a variety of ways. One intuitive conceptualization of fairness is the rule that a responder returns to the sender half of what they receive. However an alternative rule is that responders should approximate a fair division of the total amount of money available. This view predicts that responders respond to high (but not low) transfers with substantial returns – their reciprocity is contingent on what they receive. Indeed, in the current study we showed that individuals spontaneously fall into one of three categories of reciprocity: exploiter, perfect reciprocator, and contingent reciprocator. Exploiters are individuals who never fully reciprocate regardless of how much they are trusted, perfect reciprocators are individuals who always reciprocate, and contingent reciprocators are individuals who reciprocate as a function of how much they are trusted. This was demonstrated across both online and laboratory samples, and shows that the intuitive binary division of individuals into "trustworthy" or "untrustworthy" needs to be expanded.

Because our division of these different types exhaustively captured the behavior of our responders, it stands to reason that they should be recognizable by other members of these populations. Accordingly, when naïve individuals were required to learn how much to trust these different trustworthy types through trial-and-error, they were able to differentiate the different categories, however they did so suboptimally and did not profit maximally from the interactions. We showed that individuals utilized a combination of reinforcement learning (RL) and belief-based (BB) learning in a hybrid fashion, so-called experience weighted attraction (EWA). This means that subjects were not relying completely on what they themselves experienced, but also simulated the outcomes of unselected options when learning who reciprocates and who exploits. Furthermore, participants engaged more with BB learning for options that could potentially yield higher relative to lower payoffs, suggesting that participants interacting in our TG set-up were relatively optimistic and reluctant to imagine negative relative to positive outcomes. Finally, we showed that the degree to which individuals relied on BB learning was positively and robustly correlated with their average payoff, suggesting that the propensity to simulate potential outcomes serves a direct benefit to the individual.

Previous researchers have modeled trust-learning using simple reinforcement learning (Parnamets, Granwald, & Olsson, 2018), Bayesian preference learning (Devaine & Daunizeau, 2017; Hula, Montague, & Dayan, 2015), as well as EWA similar to that used in our study (Camerer & Ho, 1999; Camerer et al., 2002). However, in these studies, the decisions available to senders in the TG were severely restricted, often to a simple binary decision of trust vs. not trust (e.g. Camerer et al., 2002; Parnamets et al., 2018) which, as mentioned above, does not capture the reciprocity distributions we found in the sampled responders – particularly the contingent reciprocators found in Experiments 1 and 2. In our study, therefore, we made no such binary restriction, and allowed subjects to make any transfer between 0 and 20, which had the benefit of increasing ecological validity at the cost of a computational challenge in the form of a large option space. We met this challenge in part by estimating priors as parameters of the response function that we assumed our senders imagined responders to possess, an assumption we feel is tenable and one likewise corroborated by our model's goodness of fit as well as validity checks vis-à-vis parameter recovery. It also allowed us to gain novel insights about how participants simulated foregone payoffs. We tested different versions of our EWA model which made different predictions about how participants could estimate the payoffs of options that they did not select. By testing these different hypotheses against each other, we were able conclude that participants extrapolate from the percentage of the selected option that the responder returned. This means that senders were preferentially focusing on their own payout relative to that of the responder instead of focusing on their own total payout.

While others before us have shown that adding subjective belief to learning models improves model fits relative to simple RL (Camerer & Ho, 1999; Camerer et al., 2002; Zhu et al., 2012), these researchers did not relate reliance on belief to actual performance outcome. Here we showed that a higher propensity to engage in "fictitious play", in which the outcomes of unselected actions are mentally simulated, led to more profit than focusing solely on the actual outcome experienced. Therefore, reliance on belief, that is – imagining *potential* outcomes – not only improves model fit but affords a tangible benefit to individuals when learning to trust.

Our study leaves several open questions that future research should address. Foremost among them: what traits predict what reciprocating category an individual falls within? More research will be needed, with a more systematic battery of psychological appraisals, in order to detect what leads an individual to fall within one of these reciprocity categories. The second outstanding question demanding more research is: why are senders so far from optimality when learning these reciprocity types? While our model fits behavior well, and even helps explain inter-individual differences in reward, the current set-up does not allow us to explain what leads individuals to learn these types so poorly. Future research, potentially involving observational learning or neuroimaging, may shed light on these questions.

In sum, we show that a simple binary division of individuals into trustworthy or untrustworthy fails to capture a sizeable minority of the population (between 17% and 33%), whose reciprocity behavior is contingent on how much they are trusted. Furthermore we employ a novel application of a well validated computational model to demonstrate that individuals rely both on their experiences as well as their beliefs about unexperienced outcomes when learning how much to trust. Our results suggest that forming a mental image of the social landscape is crucial when learning who is trustworthy and who is not.

References

Balliet, D., & Van Lange, P. A. M. (2013). Trust, conflict, and cooperation: A meta-analysis. *Psychological Bulletin*, *139*(5), 1090–1112. https://doi.org/10.1037/a0030939

Baumard, N., André, J. B., & Sperber, D. (2013). A mutualistic approach to morality: The evolution of fairness by partner choice. *Behavioral and Brain Sciences*, *36*(1), 59–78. https://doi.org/10.1017/S0140525X11002202

Behrens, T. E. J., Hunt, L. T., & Rushworth, M. F. S. (2009). The Computation of Social Behavior. *Science*, *324*(5931), 1160–1164. https://doi.org/10.1126/science.1169694

Behrens, T. E. J., Hunt, L. T., Woolrich, M. W., & Rushworth, M. F. S. (2008). Associative learning of social value. *Nature*, *456*(7219), 245–249. https://doi.org/10.1038/nature07538

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

Bohnet, I., Greig, F., Herrmann, B., & Zeckhauser, R. (2008). Betrayal Aversion: Evidence from Brazil, China, Oman, Switzerland, Turkey, and the United States. *American Economic Review*, *98*(1), 294–310. https://doi.org/10.1257/aer.98.1.294

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

Camerer, C. F., Ho, T. H., & Chong, J. K. (2002). Sophisticated Experience-Weighted Attraction Learning and Strategic Teaching in Repeated Games. *Journal of Economic Theory*, *104*(1), 137–188. https://doi.org/10.1006/jeth.2002.2927

Daunizeau, J., Adam, V., & Rigoux, L. (2014). VBA: A Probabilistic Treatment of Nonlinear Models for Neurobiological and Behavioural Data. *PLoS Computational Biology*, *10*(1). https://doi.org/10.1371/journal.pcbi.1003441

Daw, N. D. (2011). Trial-by-trial data analysis using computational models. In *Decision Making, Affect, and Learning* (Vol. 6, pp. 3–38). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199600434.003.0001

De Dreu, C. K. W., & Gross, J. (2018). Revisiting the Form and Function of Conflict: Neurobiological, Psychological and Cultural Mechanisms for Attack and Defense Within and between Groups. *Behavioral and Brain Sciences*, (2019). https://doi.org/10.1017/S0140525X18002170

Devaine, M., & Daunizeau, J. (2017). Learning about and from others' prudence, impatience or laziness: The computational bases of attitude alignment. *PLoS Computational Biology*, *13*(3), 1–28. https://doi.org/10.1371/journal.pcbi.1005422

Engelmann, J. B., Schmid, B., De Dreu, C. K. W., Chumbley, J., & Fehr, E. (2019). On the psychology and economics of antisocial personality. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(26), 12781–12786. https://doi.org/10.1073/pnas.1820133116

Erev, I., & Roth, A. E. (1998). Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria. *American Economic Review*, *88*(4), 848–881. https://doi.org/10.2307/117009

Fehr, E., & Schmidt, K. M. (1999). A Theory of Fairness, Competition, and Cooperation. *The Quarterly Journal of Economics*, *114*(3), 817–868. https://doi.org/10.1162/003355399556151

FeldmanHall, O., Dunsmoor, J. E., Tompary, A., Hunter, L. E., Todorov, A., & Phelps, E. A. (2018). Stimulus generalization as a mechanism for learning to trust. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(7), E1690–E1697. https://doi.org/10.1073/pnas.1715227115

Fudenberg, D., & Levine, D. (1998). *The Theory of Learning in Games*. Cambridge, MA: The MIT Press.

Heap, S. P. H., & Zizzo, D. J. (2009). The value of groups. *American Economic Review*, *99*(1), 295–323. https://doi.org/10.1257/aer.99.1.295

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., … Tracer, D. (2005). “Economic man” in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. *Behavioral and Brain Sciences*, *28*(6), 795–815. https://doi.org/10.1017/S0140525X05000142

Ho, T. H., Camerer, C. F., & Chong, J. K. (2007). Self-tuning experience weighted attraction learning in games. *Journal of Economic Theory*, *133*(1), 177–198. https://doi.org/10.1016/j.jet.2005.12.008

Hula, A., Montague, P. R., & Dayan, P. (2015). Monte Carlo Planning Method Estimates Planning Horizons during Interactive Social Exchange. *PLoS Computational Biology*, *11*(6). https://doi.org/10.1371/journal.pcbi.1004254

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

Levy, D. J., & Glimcher, P. W. (2012). The root of all value: a neural common currency for choice. *Current Opinion in Neurobiology*, *22*(6), 1027–1038. https://doi.org/10.1016/j.conb.2012.06.001

Palminteri, S., Wyart, V., & Koechlin, E. (2017). The Importance of Falsification in Computational Cognitive Modeling. *Trends in Cognitive Sciences*, *21*(6), 425–433. https://doi.org/10.1016/j.tics.2017.03.011

Parnamets, P., Granwald, T., & Olsson, A. (2018). Building and Dismantling Trust: From Group Learning to Character Judgments, 846–851.

Prochazkova, E., Prochazkova, L., Giffin, M. R., Scholte, H. S., De Dreu, C. K. W., & Kret, M. E. (2018). Pupil mimicry promotes trust through the theory-of-mind network. *Proceedings of the National Academy of Sciences*, *115*(31), E7265–E7274. https://doi.org/10.1073/pnas.1803916115

Rand, D. G., Greene, J. D., & Nowak, M. A. (2012). Spontaneous giving and calculated greed. *Nature*, *489*(7416), 427–430. https://doi.org/10.1038/nature11467

Romano, A., Balliet, D., & Wu, J. (2017). Unbounded indirect reciprocity: Is reputation-based cooperation bounded by group membership? *Journal of Experimental Social Psychology*. https://doi.org/10.1016/j.jesp.2017.02.008

Romano, A., Balliet, D., Yamagishi, T., & Liu, J. H. (2017). Parochial trust and cooperation across 17 societies. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1712921114

Rutledge, R. B., Dean, M., Caplin, A., & Glimcher, P. W. (2010). Testing the reward prediction error hypothesis with an axiomatic model. *Journal of Neuroscience*, *30*(40), 13525–13536. https://doi.org/10.1523/JNEUROSCI.1747-10.2010

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement Learning, Second Edition: An Introduction*. *The MIT Press*.

Zhu, L., Mathewson, K. E., & Hsu, M. (2012). Dissociable neural representations of reinforcement and belief prediction errors underlie strategic learning. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(5), 1419–1424. https://doi.org/10.1073/pnas.1116783109

4

Chapter 4

Generosity Biases the Learning of Cultural Conventions

Michael Rojek-Giffin[1,2*], Maël Lebreton[3,4,*, #], Jean Daunizeau[5,6,7], Andrea Fariña[1,2], Jörg Gross[1,2], and Carsten K.W. De Dreu[1,2,8,#]

1 Institute for Psychology, Leiden University, Leiden, the Netherlands
2 Leiden Institute for Brain and Cognition, Leiden University, Leiden, the Netherlands
3 Laboratory for Behavioral Neurology and Imaging of Cognition, Department of Basic Neurosciences, Université de Genève, Geneva, Switzerland
4 Swiss Centre for Affective Sciences, Université de Genève, Geneva, Switzerland
5 Sorbonne Université, Paris, France
6 Institut du Cerveau et de la Moelle épinière, Paris, France,
7 INSERM UMRS 1127, Paris, France
8 Center for Research in Experimental Economics and Political Decision Making, University of Amsterdam, Amsterdam, the Netherlands
* Indicates shared first authorship
# Corresponding authors at Leiden University, PO 9555, 2300 RB Leiden, the Netherlands; email: c.k.w.de.dreu@fsw.leidenuniv.nl and Geneva University, Geneva, Switzerland; email: mael.lebreton@unige.ch

Summary

Human groups can markedly differ in fairness and cooperation norms, and these differences can create intergroup misunderstandings and conflict. At the same time, humans also trade and travel across cultural divides, suggesting that they can learn and adapt to new culture-specific conventions and rules of engagement. While such adaptions avoid intergroup conflict and benefit intergroup exchange, how humans learn group-specific rules that are often implicit and distinct from already learned values and norms remains poorly understood. Here we examine this fundamental learning process underlying social rule acquisition. We created three populations with different yet unobservable rules of engagement and varied whether or not decisions affected interaction partner outcomes. Participants made bargaining offers to responders from these different populations and could observe whether their offer was accepted or rejected. Participants quickly adapted to group-specific rules in learning environments without social consequences, but were overly generous and ended up misrepresenting what would be acceptable when decisions affected their partner's outcomes. We propose a computational model, combining Bayesian principles and social preferences, that mechanistically explains how generosity leads to biased sampling, impeded learning, and false beliefs about what offers are deemed acceptable. Using functional neuroimaging, we mapped key computational variables in two major brain networks, previously associated with value-based and social decision-making. Results suggest that generosity, related to brain regions associated with decision-conflict and perspective-taking, can induce self-fulfilling beliefs in pro-sociality norms that may help to increase cooperation and reduce conflict between distinct groups but also create inaccurate stereotypes and economic inefficiencies.

Key Words: Decision Neuroscience | Moral Sentiments | Bayesian Learning | Social Norms

Main Text

The myriad agreements that govern social life require some form of coordinated negotiation about what to give and what to take, and what to do and to avoid (Crawford, 2019; De Dreu, Weingart, & Kwon, 2000; King-Casas et al., 2005). Social preferences for fairness and the welfare of other, alongside norms for cooperation, can help people to avoid economically costly coordination failures and emotionally taxing impasses that emerge when offering too little or asking too much (De Dreu, Gross, Fariña, & Ma, 2020; Fehr & Fischbacher, 2002; Fehr & Schmidt, 1999). The coordination value of social preferences and cooperation norms decreases, however, when partners differ in their rules of engagement and hold different conceptions of fairness (Bicchieri, 2005; Fiske, 1992; Sam & Berry, 2010). Indeed, fairness norms and rules of engagement differ across groups, markets, and cultures—what is considered fair and appropriate in some cultures may be considered offending and too demanding in others (Blake et al., 2015; Debove, Baumard, & André, 2016; Henrich et al., 2005).

While operating on one's fairness considerations may facilitate coordination and social exchange within cultures and markets (Bicchieri, 2005; Fehr & Fischbacher, 2002; Tomasello, Carpenter, Call, Behne, & Moll, 2005), across cultures it can increase rather than decrease misunderstanding and conflict (Adair, Okumura, & Brett, 2001; Gelfand, Erez, & Aycan, 2007; Gelfand & Harrington, 2015; Rai & Fiske, 2011; Weber & Camerer, 2003). Extant work on mutual gains bargaining has indeed shown that culture-specific norms of fairness facilitate intracultural deal-making yet undermine collective efficiency when participants come from markedly different cultures (Buchan, Croson, & Johnson, 2004; Liu, Huang, Luo, & Zhao, 2012). And yet, since prehistoric times, humans interact and trade across cultural boundaries (Rand & Nowak, 2013; Soares et al., 2010), suggesting an ability to learn and adapt to new culture-specific rules of engagement. Here we address this possibility, asking how humans learn new rules of engagement, and to what extent culturally engrained social concerns for fairness modulate such learning and adaptation in intercultural interactions and economic exchanges.

Creating cultural conventions in the lab

In a first step, we created three "cultures" in the lab that differed in the extent to which individual members would accept (versus reject) ultimatum offers. Participants (Experiment 1, N = 210, see *Methods*) decided to accept or reject a range of possible offers from proposers (out of endowment e = 20). When accepting the offer, participants as responders earned the offer and their proposer earned e – offer. When rejecting, each side earned 0 (Güth, Schmittberger, & Schwarze, 1982). In this simple game, modeling a social transaction between two people, the proposer has to make an offer large enough to reduce the risk of rejection without knowing what the responder deems appropriate or acceptable a priori. If the proposer offers too much, she runs the risk of being overly generous, losing money that she could have kept for herself. Whereas proposers always had $e = 20$, responders indicated their accept/reject decisions when they themselves had $e = 0$, $e = 10$, and $e = 20$ already in their pocket to induce and create different acceptance thresholds (*Methods*). Confirming that different responder endowments influenced rules of engagement (or acceptance thresholds; (Slonim & Roth, 1998)), we find that responders with $e = 0$ exhibited an acceptance distribution commonly found in Western societies (Henrich et al., 2005; Oosterbeek, Sloof, & van de Kuilen, 2004). Thresholds were significantly lowered when responders' had a starting endowment of $e = 10$ and $e = 20$, respectively (**Fig 1A**), corresponding to rules of engagement found in some non-Western cultures (Blake et al., 2015; Henrich et al., 2005). Different responder endowments thus created different group-specific rules of engagement (i.e. what was deemed a "fair" offer that was accepted rather than rejected), which we modelled with sigmoid functions - thereafter referred to as acceptance functions. These sigmoid functions formally describe the probability of each offer being accepted in each culture, and are fully characterized by two parameters (an intercept and a slope - **Fig 1B**).

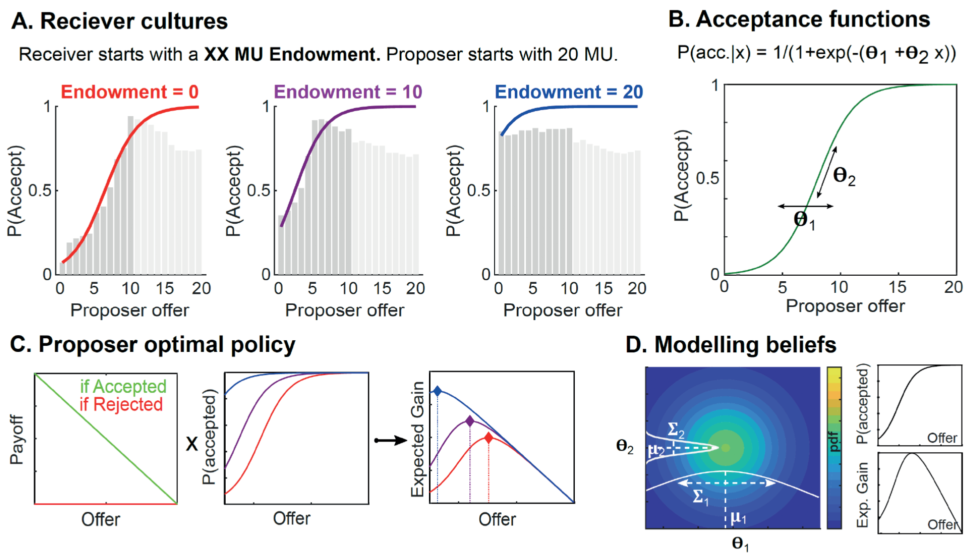

Figure 1.
A model to learn cultural conventions. **(A) Group-specific Rules of Engagement**, estimated from Experiment 1. Grey histograms represent acceptance frequencies for respondents with starting endowments of (from left to right) 0, 10, and 20 MU, to any possible offer from a proposer with 20MU starting endowment. Colored lines represent acceptance functions, i.e. logistic functions fitted to the data to characterize the entire population. **(B) Acceptance functions**. Acceptance functions are modelled as logistic functions, fully characterized by two parameters: an intercept θ1 and a slope θ2.

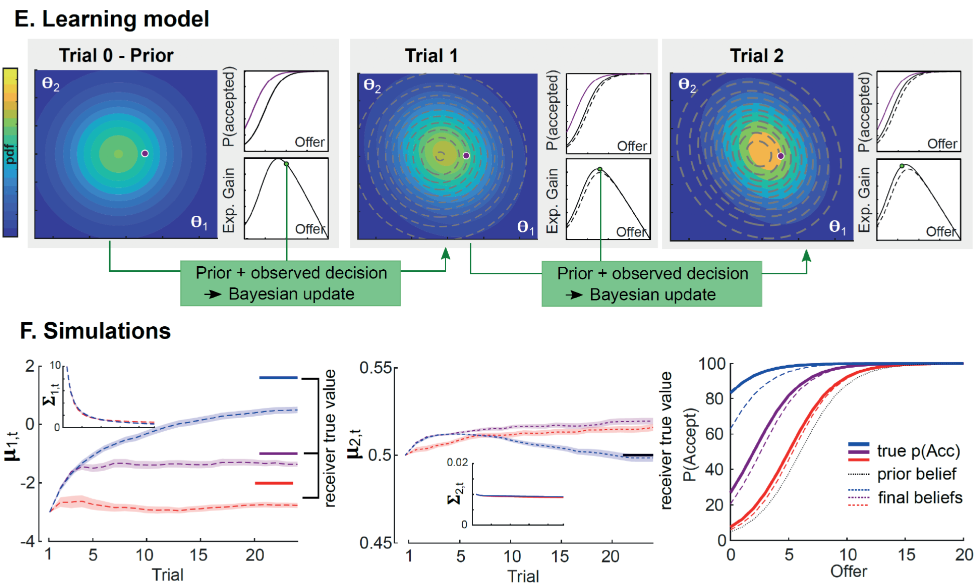

Figure 1. Continued.
(C) Proposer optimal policy. The ultimatum payoff structure (left panel) is combined with (known) acceptance functions (middle panel) to derive expected gain, as a function of offers, for each culture (right panel). The optimal policy is to select the offer with the maximum expected gain (diamonds). **(D) Modeling beliefs**. Proposer typically do not know the true acceptance function parameters, but are endowed an internal representation – belief – of those, which take a Gaussian form $p(\theta) = N(\mu,\Sigma)$. The colored surface on the right panel represent the belief multivariate probability density function (with the marginal probability distribution for each parameter represented as white curves). **(E) Learning model**. Consider a proposer, represented by her belief probability density function (colored surface), confronted with a specific culture, represented by the parameters of its acceptance function (purple dot). At each trial, the proposer uses her estimated acceptance function (black curve; top right insets) to produce an expected gain function (bottom right insets), and select an offer that (soft) maximize expected gain (green dot). In this case, the offers are accepted, and the proposer use this information to update her beliefs using Bayes rule. Note that the peak of the belief probability density function gets closer to the true parameters (purple dots). Dotted lines represent the previous trial features. **(F) Simulations**. Simulations (N = 100) show that the Bayesian scheme used can efficiently converge to good approximation of the intercept (left panel) and slope (middle panel) of different culture's acceptance functions (color codes are identical to panel A). Right panel pictures the original (dotted black line) and final estimated acceptance functions (dotted colored lines), with true acceptance function superimposed (thick colored lines).

A model of learning cultural conventions

We next developed a mathematical framework to understand how ideal proposers would behave when repeatedly paired with responders sampled from those different cultures. First, consider an agent whose sole goal is to maximize her monetary gains. Should the agent know the acceptance function parameters of the different cultures, her optimal strategy is to use this function to compute each offer's expected gain, knowing the ultimatum offer payoff rule, and to select the offer that maximizes her expected payoff (**Fig 1C**). Yet, in ecological settings, an *uninformed* agent does not know the acceptance function's parameters characterizing each culture (FeldmanHall & Shenhav, 2019). Instead, she relies on beliefs about these parameters (**Fig 1D**). Using Bayesian learning principles (see *Materials and Methods*), those (prior) beliefs can be updated when observing responders' decisions to offers made, to form posterior beliefs that better estimate the true parameters (**Fig 1E**). Simulations demonstrate that an approximate implementation (based on a variational-Laplace scheme) of this optimal Bayesian learning principle can efficiently learn different culture acceptance function parameters, while selecting offers that (soft-)maximize her expected payoff (**Fig 1F**).

A mechanistic explanation of how generosity biases cultural learning

Now consider an agent who is not only concerned about maximizing her gains, but also by the responder's welfare. Such social concern can be expressed, for instance, as inequality aversion (see also *Materials and Methods*). Then, should the agent know the different acceptance functions, her optimal policy is nonetheless modified, because the value derived from an offer now integrates both an expected gain term, and a term that accounts for social concern. To fully satisfy her preferences, this socially-concerned agent should make higher offers than a similar but purely gain maximizing agent (**Fig 2A**). A socially-concerned *uninformed* proposer can also leverage the Bayesian update rule to learn cultural conventions. Yet, simulations show that social concerns bias the estimation of the acceptance function parameters. More specifically, despite the optimal Bayesian update rule, socially-concerned learners end up misrepresenting low-acceptance threshold cultures, believing that responders require higher offers than necessary (**Fig 2B**). This result can be explained under the framework of efficient hypothesis testing: conditional on the agent's current beliefs, offers (and responders' decisions to those offers) are not all equally *efficient* (i.e. informative or diagnostic; see *Materials and Methods*). Social concerns bias offers upward, away from the offers for which reactions would be the most efficient to adjust beliefs to the true acceptance functions, effectively biasing posterior beliefs (**Fig 2C**).

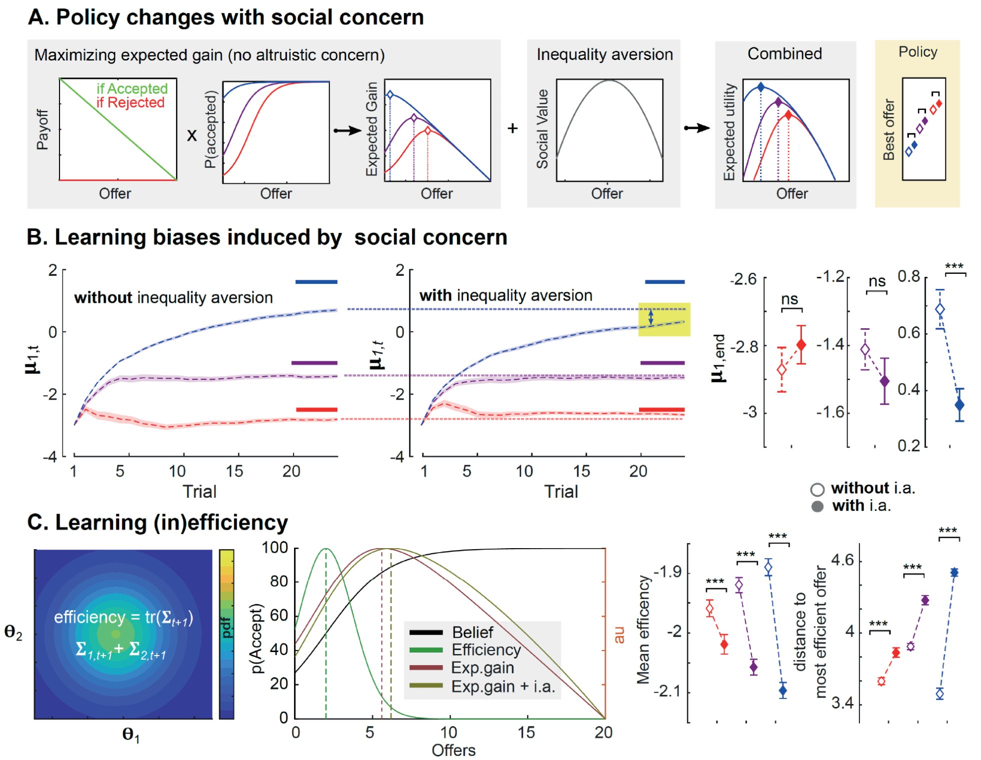

Figure 2. How Generosity Biases the Learning of Cultural Conventions. **(A) Proposer optimal policy**. The ultimatum payoff structure is combined with acceptance functions to derive expected gain, as a function of offers, for each culture (left panel). An inequality aversion (IA) term (middle panel) can be added to generate a complex expected utility function (right panel). The new optimal policy is to select the offer with the maximum expected utility (diamonds). The addition of the IA term therefore leads to higher optimal offers (right-most panel: plain diamonds: with IA.; empty diamonds: without IA). **(B) Simulations**. Two sets of simulations (N = 100 each) were performed with (middle panel) or without (left panel) the inclusion of IA term to select the offer. Results show that after 24 trials, Bayesian learning converged to different beliefs about the acceptance function intercepts (μ1) in those two conditions, especially in the most accepting culture (blue color – right panel). **(C) Learning in(efficiency)**.
*** P < 0.001

An experimental framework to test the culture learning model

To test model predictions, we performed three experiments. Participants in the role of proposers (N = 198) made offers to responders that were identified by three neutral symbols, similar to culture-specific identity markers such as language or clothing (**Fig 3A**). Unbeknownst to proposers, symbols corresponded to a particular acceptance threshold established in Experiment 1. Through a process of offer approval and rejection, proposers could learn and adapt to these group-specific rules of engagement, akin to learning new norms and culture-specific conventions (Sam & Berry, 2010). In addition, in one treatment (henceforth social condition), proposers interacted with human responders whose earnings depended on the (acceptance of) proposed offers whereas in another treatment (henceforth non-social condition), proposers interacted with behaviorally identical computer agents that did not earn from the (acceptance of the) offer (Baumgartner, Fischbacher, Feierabend, Lutz, & Fehr, 2009; Sanfey, Rilling, Aronson, Nystrom, & Cohen, 2003). In the social condition, participants were explicitly told that they were facing groups of responders who had received different starting endowments but not what the endowments were. In the non-social condition, participants were told that they facing computer generated lotteries programmed to mimic the behavior of participants who had received different starting endowments (see *Materials and Methods*). This created a non-social and social learning environment with identical learning contingencies, elucidating how social concerns for the responders affects learning (see also *Materials and Methods*).

Social concerns modulate learning behavior

Proposers indeed behaved differently when their decisions had social consequences. Proposers facing human responders made higher initial offers compared to computer responders (**Fig 3BC**: $b \pm$ se = 0.852 ± 0.231, $p < 0.001$) and higher offers on average ($b \pm$ se = 0.570 ± 0.026, $p < 0.001$). In sum, proposers were more generous to human as opposed to computer opponents. As a consequence, earnings were significantly lower when playing against human rather than computer responders (**Fig 3C:** $b \pm$ se = -0.185 ± 0.055, $p < 0.001$).

Results also showed that proposers progressively learn the group-specific rules of engagement across repeated offers to responders. Offers converged over trials on three different final offers in the three different conditions (**Fig 3BC**; $b \pm$ se = -1.718 ± 0.076, $p < 0.001$). After learning, offers were still significantly higher for humans than for computer responders (**Fig 3BC;** $b \pm$ se = 0.596 ± 0.146, $p < 0.001$).

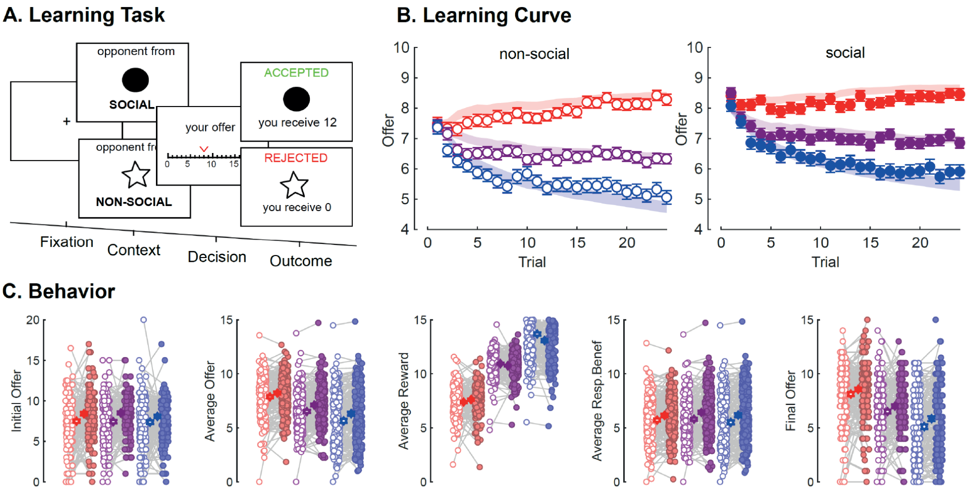

Figure 3. Behavioral results of learning task. **(A) Trial timeline**. Subjects made ultimatum offers to three different responder groups each marked with a neutral shape in alternating blocks either with or without social consequences. **(B) Offers over trials**. Across multiple encounters, offers converge on the acceptance thresholds of the three different responder populations, depicted with different colors. Dots with error bars represent mean ± standard error, shaded areas represent model predictions. Social preferences (human versus computer-simulated responders) impede convergence of offers on responder acceptance thresholds (shown m ± se). **(C) Behavioral results**. Behavior differed between responder group and social consequence in terms of initial offer, average offer, average reward, average responder benefit, and final offer.

Social concerns bias posterior beliefs

There may be two mechanisms accounting for the higher offers to human than computer responders at the end of learning. First, proposers may converge to similar estimates of acceptance functions in both social and non-social conditions, but still make higher offers to human responders because of social concerns. Our Bayesian model suggests, however, that by making higher offers to human rather than computer responders, proposers converge to different estimates of (the same) acceptance functions in the social versus non-social condition. Whereas the second possibility presumes posterior beliefs to be biased by the social versus non-social condition, the first possibility does not. To examine posterior beliefs, participants in two of our experiments (N = 93) performed, after the learning task, an incentivized belief estimation task that directly elicited participant beliefs about the acceptance functions (*Materials and Methods*). We find that, as predicted by the Bayesian model, proposers estimated the acceptance thresholds of human responders to be higher than those of computer-simulated responders, especially for the most lenient responder group ($t(92) = 1.914$, $p = 0.059$, **Fig. 4BC**). This indicates that social concerns can bias the formation and updating of group-specific beliefs about rules of engagement.

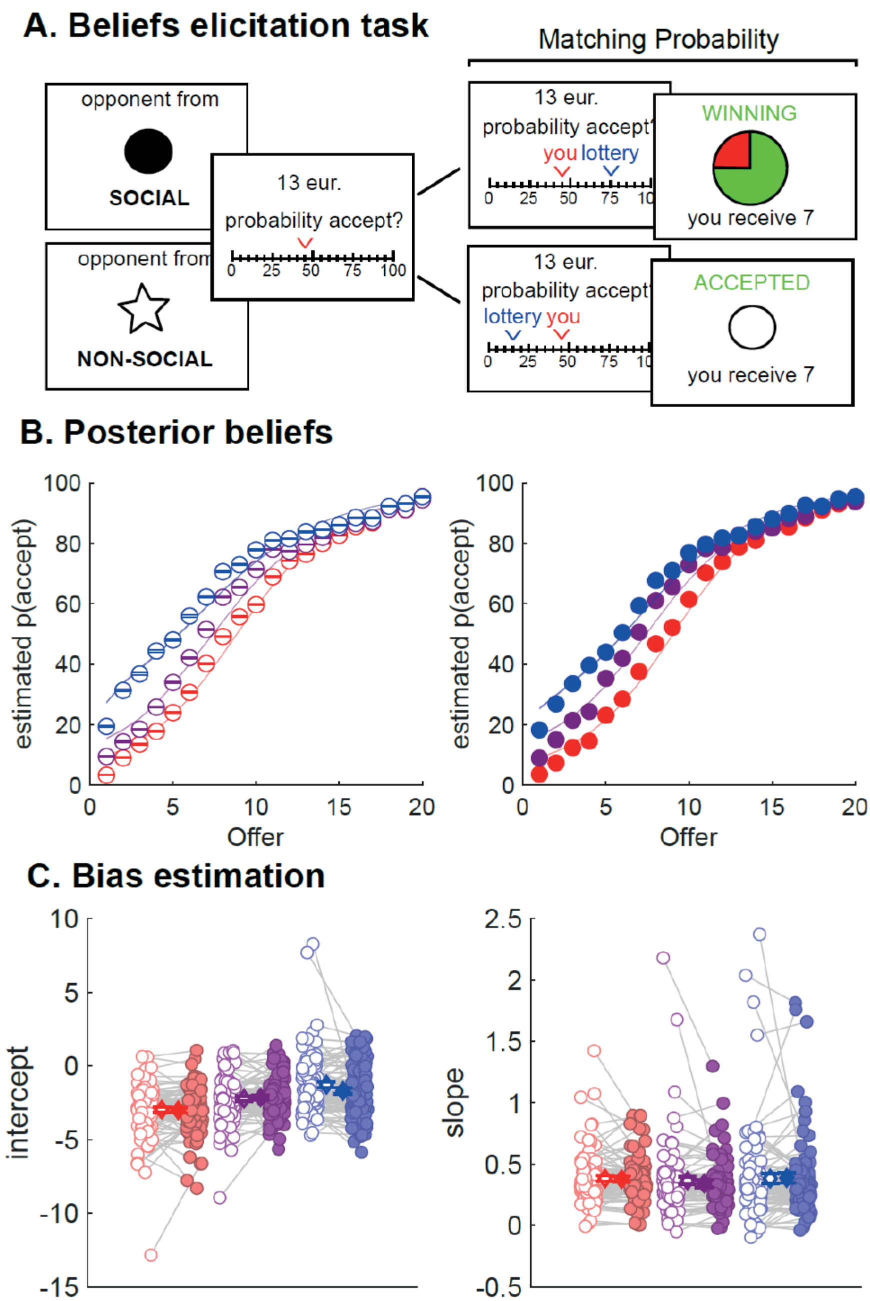

Figure 4. Social preferences modulate beliefs. **(A) Trial timeline**. Proposers were asked what probability each offer had of being accepted by each responder type in both social and non-social conditions. **(B-C) Posterior beliefs**. Proposer answers were averaged (dots with error-bars) and fitted to sigmoid functions (lines), which represent proposers' estimates of each responder's acceptance function. The intercepts and slopes were then averaged within each responder and tested for differences between the social and non-social conditions.

Fitting Bayesian learning with inequality aversion captures human behavior

Participants' behavior match predictions from our computational model simulations, but this in itself is no evidence that our model can truly and most parsimoniously explain the (social-concern biased) learning and updating of culture-specific rules of engagement (Palminteri, Wyart, & Koechlin, 2017). However, a set of model fitting and model comparisons exercises (Daw, 2011) showed that the simulations of the estimated parameters from the Bayesian learning with social concerns model mimicked the key behavioral patterns observed in our participants: learning curves per condition (**Fig. 3B**) and average behavior (**Fig. 3C**). In addition, the model posterior beliefs estimated at the end of learning clearly accounted for the social bias in the posterior belief task (**Fig. 4C**). A Bayesian Model Comparison between Bayesian learning models with and without social concerns revealed that models with social concerns better account of the data than models without social concern. Combined, these results indicate that our model provides a satisfactory mechanistic explanation of participant behavior.

Functional neuroimaging

Neuroimaging data (see *Materials and Methods* revealed for details on methods) revealed that proposer brains exhibited different patterns of neural activity depending on whether or not their decisions had social consequences. A whole-brain searchlight procedure using multi-voxel pattern analysis (*Materials and Methods*) revealed significant neural differences between offers made to human rather than computer responders in the dorsal anterior cingulate cortex (dACC) and the precuneus during feedback (**Fig. 5A**), and in dACC and in the right superior temporal sulcus (STS) when participants made their offers (**Fig. 6B**). This suggests that especially the dACC was critically involved in the differential behavior and beliefs observed between human and computer-simulated opponents.

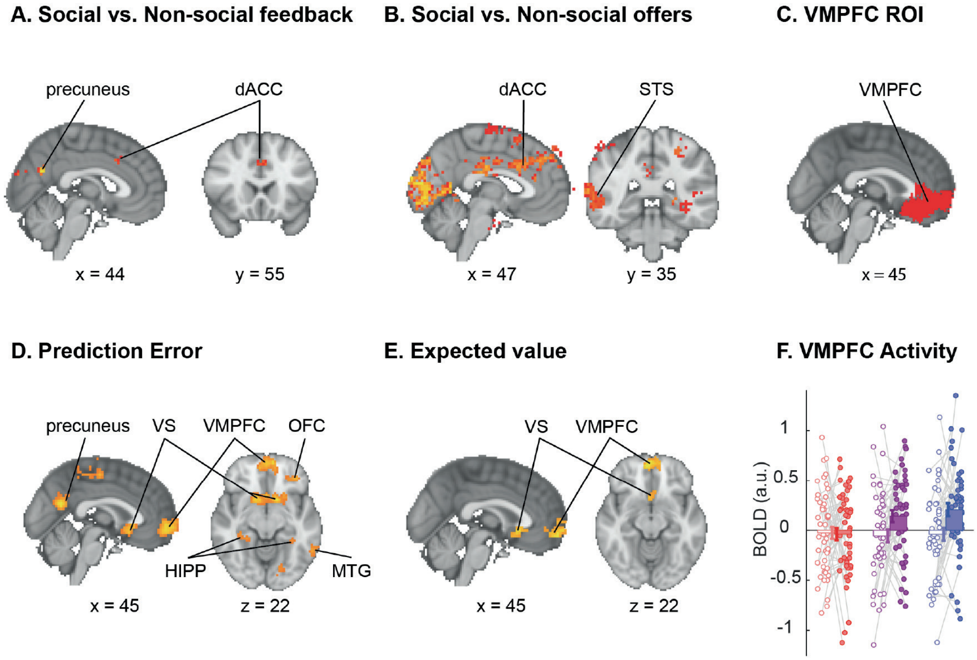

Figure 5. Neural activity differs between conditions and is tracked by model parameters. **(A-B) Social context in the brain**. Proposer brains exhibited significantly different patterns of neural activation during both feedback **(A)** and while making offers **(B)**. **(C) Region of interest**. A priori defined ventromedial prefrontal cortex (VMPFC) ROI obtained from meta-analytic tool Neurosynth (Yarkoni, Poldrack, Nichols, Van Essen, & Wager, 2011). **(D-E) Neural correlates of model parameters**. Regions encompassing value attribution and social cognition significantly correlated with prediction errors (**D**) and expected values (**E**) from our Bayesian Preference Learner model. **(F) VMPFC Social × Responder group interaction**. Neural activity in the VMPFC exhibited an interaction between social context and responder group, with the most lenient responder group (blue) eliciting a higher VMPFC BOLD response in the social relative to the non-social context, while this social/non-social differentiation was absent in the most stringent group (red).

Results also showed key parameters of our Bayesian learning model to covary with valuation and reward learning neural circuitry. The Bayesian model's prediction error significantly correlated with BOLD response during feedback in the ventral striatum (VS) and ventromedial prefrontal cortex (VMPFC), as well as the posterior cingulate cortex (PCC), orbitofrontal cortex (OFC), hippocampus, and STS (**Fig. 5D**). The Bayesian model's expected value significantly correlated with BOLD response during offer selection in the VS and VMPFC (**Fig. 5E**).

When actions result in a social consequence, both simulations and behavioral results demonstrate that the most lenient responder group has their acceptance thresholds under-estimated to a higher degree than the other responder groups, and hence are the most over-valued (**Fig 2B & Fig 3BC**). This same interaction between responder groups and social context was also expressed neurally. The correlation between VMPFC BOLD response and the Bayesian learning model's expected value significantly differed between responder groups depending on social context, with the difference being most pronounced for the most lenient responder group (social ´ responder (e = 10): b ± se = 0.209 ± 0.123, p = 0.091; social ´ responder group (e = 20): b ± se = 0.259 ± 0.123, p = 0.037; social responder group (e = 0) vs. non-social responder group (e = 0): $t(48)$ = -0.136, p = 0.892; social responder group (e = 10) vs. non-social responder group (e = 10): $t(48)$ = 2.259, p = 0.028; social responder group (e = 20) vs. non-social responder group (e = 20): $t(48)$ = 2.888, p = 0.006; **Fig. 5F**). This indicates that the process of learning culture-specific rules of engagement is hindered by social concern in the form of inequity aversion, and this hinderance is evident at the computational, behavioral, and neural level.

Conclusions

By experimentally isolating fairness norms and concomitant rules of engagement as a core component along which human cultures differ (Blake et al., 2015; Fiske, 1992; Gelfand & Harrington, 2015; Henrich et al., 2005; Oosterbeek et al., 2004; Rai & Fiske, 2011), we investigated how humans learn implicit, culture-specific conventions in a controlled learning environment. Specifically, participants learned to match their ultimatum offers to unobservable responders' acceptance thresholds and updated their behavior and beliefs about culture-specific rules of engagement accordingly. At the same time, participants exhibited generosity when facing human opponents which impeded their ability to learn. Especially with responders with lenient acceptance thresholds, people learned slowly and continued to make unnecessarily high offers, often not observing that more self-serving choices would have been deemed appropriate in some environments. This was not the case in a non-social learning environment. It thus appears that social concerns can initiate a self-fulfilling process of selective sampling and updating, whereby some culture-specific rules of engagement are better learned and adopted than others.

The self-fulfilling bias in learning revealed here provides a mechanistic explanation for asymmetric cultural evolution, in which over time some cultural conventions (Fiske, 1992), and value-systems more generally (Rai & Fiske, 2011), become more wide-spread than others. In our case, people were too generous with some partners who consistently received more than their acceptance thresholds. Because overly generous offers were frequently accepted, people were confirmed in their beliefs that others' acceptance thresholds were higher than they actually were. It stands to reason that those who persistently receive more than they need likewise, over time, update their expectations and increase their acceptance thresholds, leading to a consensually shared social convention of what is needed and required to coordinate agreements, a self-fulfilling prophecy based on generosity-impeded learning. It stands to reason too that such biased updating of beliefs and expectations operate in other culture-specific norms and conventions, explaining why some (detrimental) cultural norms and conventions survive and increase in popularity, while others fade and become extinct.

Materials and Methods

Ethics statement

Experiments received ethics approval from Leiden University (CEP17-0829/274, CEP17-1012/341, CEP19-0108/7, and CEP19-0617/350). Data were collected over four independent samples, and subsequently pooled together for analysis: Sample 1: N = 210; Sample 2: N = 50; Sample 3: N = 44; Sample 4: N = 49. Participants were recruited from the subject pool at Leiden University, provided written informed consent and were debriefed and paid for participation. Experiments were incentivized and did not involve deception. Individual anonymity was guaranteed throughout and earnings were paid in private.

Responders' tasks and data

To obtain the responder populations against which our proposers played, we first invited 210 participants to play Ultimatum Game (UG) as responders using the strategy method. Responders played in five different conditions consisting of starting endowments with different amounts of monetary units (MU): 0MU, 5MU, 10MU, 15MU or 20MU (**Fig. 6**). These different starting endowments given to the responders resulted in different amounts of money being at stake on any given trial, and effectively modeled different environments with different cultural customs. For example, when the responder received a starting endowment of 0MU, all the money at stake on that trial would be in the hands of the proposer. Therefore, a 50/50 division of the total amount at stake on such a trial would be 10MU from the proposer. However, on a trial in which the responder received a starting endowment of 10MU, the total amount at stake would now be the proposer's starting endowment (20MU) plus the responder's starting endowment (10MU), totaling 30MU at stake altogether, and making a 50/50 division equal to 5MU instead of 10MU. On trials in which the responder received a starting endowment of 20MU, there was 40MU at stake altogether, and an offer of 0MU resulted in a 50/50 division.

Responders were asked if they would accept all offers between 0 and 20 (**Fig. 6**), and their answers were then pooled and summed to obtain the frequency with which each offer was accepted in a given condition (**Fig. 6**). The condition in which responders received a starting endowment of 0MU exhibited the distribution commonly found in the Ultimatum Game, with offers of 10MU being accepted in most cases, but offers below 10MU being frequently rejected. More importantly, the conditions in which responders received initial starting endowments resulted in responders accepting lower offers. The resulting distributions of this manipulation acted as the populations against which our proposer played. Specifically, we fit logistic functions over the resulting distributions of this manipulation in order to obtain the acceptance threshold of the entire population. These functions were used to provide feedback to our proposers.

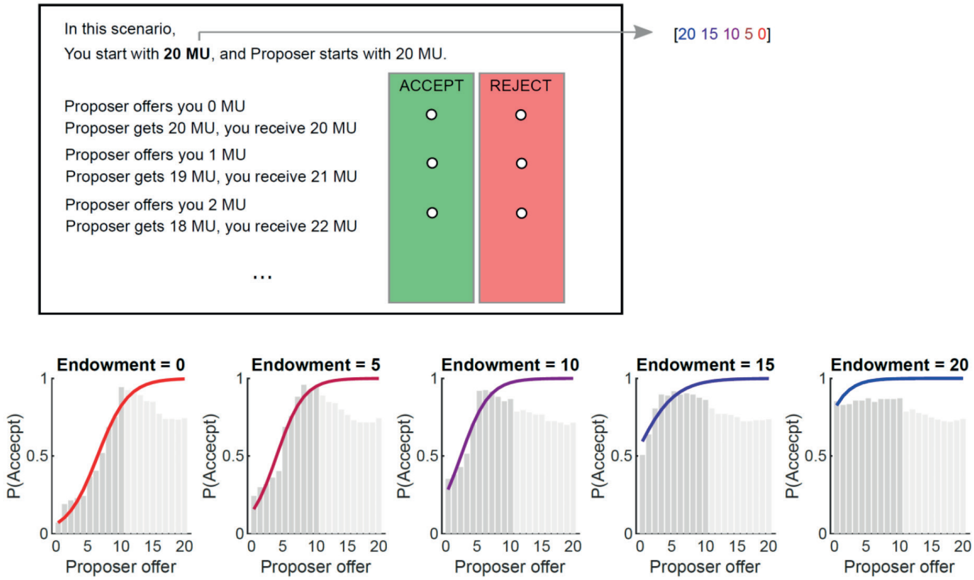

Figure 6. Responder task and data. Participants provided ultimatum game responses for each possible offer in five different conditions in which they were given five different starting endowments unaffected by the offers from proposers (top row). These starting endowments resulted in different distributions of acceptance frequencies (bottom row). To obtain the acceptance function of each responder group (i.e. each endowment), we fit logistic functions the distributions. The feedback provided to proposers was determined by the response functions for endowments of 0, 10, and 20.

Proposer Tasks

Learning task

Participants played between two blocks (Pilot: 1 human and 1 computer condition) and four blocks (fMRI and Replication: 2 human and 2 computer conditions) of the Ultimatum Game as proposers. In each block, participants played 72 trials against 3 different responder cultures with different acceptance thresholds (24 trials per responder culture). Each responder group was marked with a neutral shape such as a circle or square, and all shapes were randomized for each participant and only used once such that each block consisted of completely novel shapes. Participants were instructed that they were playing against groups of responders who had received different starting endowments, although they were not told what the endowments were. Importantly, while in the human condition participants were told that they were playing against groups of responders who had received different starting endowments, in the computer condition participants were told that they were playing against computer generated lotteries programmed to mimic the behavior of participants who had received different starting endowments. In other words, they were told that they were playing against computers programmed to behave like humans. One trial from each block was selected at random for payment.

The Pilot sample was programmed in oTree (Chen, Schonger, & Wickens, 2016) and was completely self-paced. On every trial, participants were shown a screen with a shape denoting the responder group, and selected an offer between 0 and 20. Participants were then presented with a results page containing feedback regarding whether or not their offer was accepted, and how much MU they earned for that trial. The fMRI and Replication experiments were programmed in the Psychtoolbox library of Matlab (Mathworks). Each trial started with a fixation cross (1.5 – 2.5 seconds), followed by a screen showing the shape denoting the responder culture (2 – 3 seconds). Subjects then used a slider to select an offer between 0 and 20, after which they were shown a screen indicating whether or not the offer was selected and how much MU they received for that trial (2 – 3 seconds).

Posterior belief task

After one human and one computer block (for the fMRI and Replication samples), participants completed a fully incentivized belief estimation task. Due to technical issues, three participants from the fMRI sample were unable to complete the task, leaving a final sample of N = 93. In this task subjects were asked to estimate the probability each offer had of being accepted by each responder culture against whom they had just played. On each trial, participants were presented with a shape corresponding to one of the responder cultures from the previous blocks as well as an offer between 0 and 20. They were asked to identify on a scale from 0% to 100% how likely the given offer was to be accepted by that particular responder culture. All trials were self-paced. We used a Becker-DeGroot Marschak auction to incentivize accuracy, and selected one trial at random for payment.

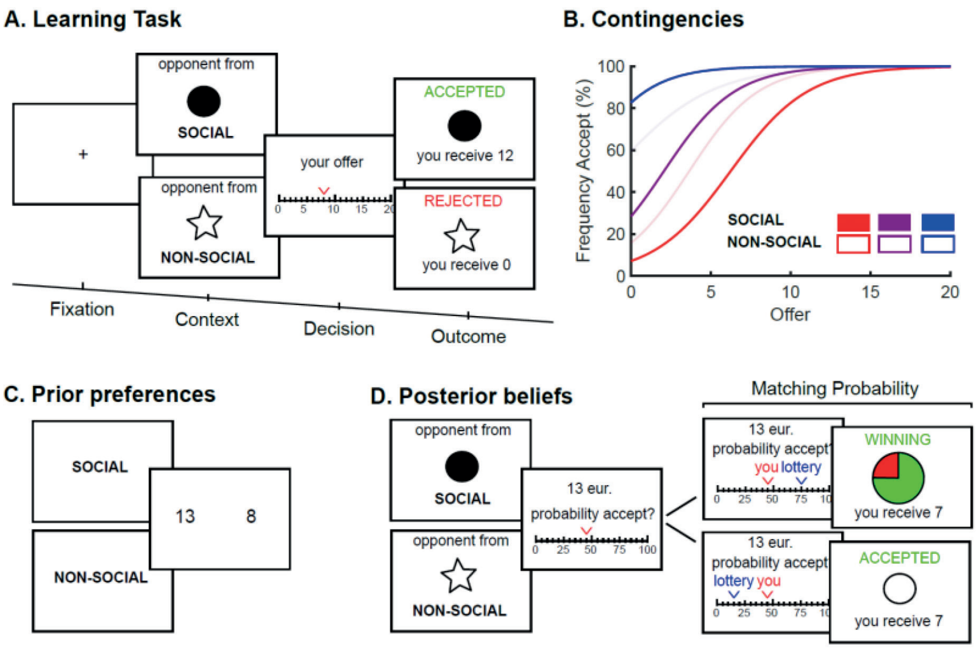

Figure 7. Schematic of all tasks. Participants completed a learning task **(A)** against three responder groups **(B)** in both social and non-social contexts. After the learning phase, subjects then completed a probability matching task **(D)** assessing their posterior beliefs regarding the acceptance functions of the different responder groups.

Neuroimaging

acquisition details

Neuroimaging was performed using a standard whole-head coil on a 3-T Philips Achieva MRI system at the Leiden University Medical Center. Participants completed four runs, during which 400 T2*-weighted whole-brain echo-planar images (EPIs) were collected (TR = 2.2 s; TE = 30 ms, flip angle = 80°, 38 transverse slices, 2.75 × 2.75 × 2.75 mm +10% interslice gap). The first five dummy scans were discarded to allow for equilibration of T1 saturation effects After each functional fun, a B0 field map was acquired. Additionally, a 3-D T1-weighted scan was acquired (TR = 9.8 ms; TE = 4.6 ms, flip angle = 8°, 140 slices, 1.166 × 1.166 × 1.2 mm, FOV = 224.000 × 177.333 × 168.000).

Neuroimaging preprocessing

fMRI data were preprocessed using FMRIPREP version 1.0.8 (Esteban et al., 2019), a Nipype (Gorgolewski et al., 2011) based tool. Each T1w (T1-weighted) volume was corrected for INU (intensity non-uniformity) using N4BiasFieldCorrection v2.1.0 (Tustison et al., 2010) and skull-stripped using antsBrainExtraction.sh v2.1.0 (using the OASIS template). Brain surfaces were reconstructed using recon-all from FreeSurfer v6.0.1 (Dale, Fischl, & Sereno, 1999), and the brain mask estimated previously was refined with a custom variation of the method to reconcile ANTs-derived and FreeSurfer-derived segmentations of the cortical gray-matter of Mindboggle (Klein et al., 2017). Spatial normalization to the ICBM 152 Nonlinear Asymmetrical template version 2009c (Fonov, Evans, McKinstry, Almli, & Collins, 2009) was performed through nonlinear registration with the antsRegistration tool of ANTs v2.1.0 (Avants, Epstein, Grossman, & Gee, 2008), using brain-extracted versions of both T1w volume and template. Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM) and gray-matter (GM) was performed on the brain-extracted T1w using fast (FSL v5.0.9) (Zhang, Brady, & Smith, 2001).

Functional data was motion corrected using mcflirt (FSL v5.0.9) (Jenkinson et al., 2002). This was followed by co-registration to the corresponding T1w using boundary-based registration (Greve & Fischl, 2009) with 9 degrees of freedom, using bbregister (FreeSurfer v6.0.1). Motion correcting transformations, BOLD-to-T1w transformation and T1w-to-template (MNI) warp were concatenated and applied in a single step using antsApplyTransforms (ANTs v2.1.0) using Lanczos interpolation.

Physiological noise regressors were extracted applying CompCor (Behzadi, Restom, Liau, & Liu, 2007). Principal components were estimated for the two CompCor variants: temporal (tCompCor) and anatomical (aCompCor). A mask to exclude signal with cortical origin was obtained by eroding the brain mask, ensuring it only contained subcortical structures. Six tCompCor components were then calculated including only the top 5% variable voxels within that subcortical mask. For aCompCor, six components were calculated within the intersection of the subcortical mask and the union of CSF and WM masks calculated in T1w space, after their projection to the native space of each functional run. Frame-wise displacement (Power et al., 2014) was calculated for each functional run using the implementation of Nipype.

Many internal operations of FMRIPREP use Nilearn (Abraham et al., 2014), principally within the BOLD-processing workflow. For more details of the pipeline see: http://fmriprep.readthedocs.io/en/latest/workflows.html.

Univariate analysis

Preprocessed functional data was then analyzed with FSL (Oxford Centre for Functional MRI of the Brain (FMRIB) Software Library; www.fmrib.ox.ac.uk/fsl). At the first level (within subjects within runs), each subjects blood oxygen level dependent (BOLD) data was spatially smoothed with 5mm FWHM gaussian kernel, high pass temporal filtered, film pre-whitened, and convolved with the canonical double gamma hemodynamic response function. The general linear model (GLM) included the following regressors: “opponent”: the time-phase when the shape representing the different responder population was presented; “expected value”: the “opponent” regressor modulated by the intercept of our RL model, and orthogonalized with respect to "opponent"; "feedback": the time-phase when the response of the opponent was presented to the subject; "prediction error": "feedback" modulated by the prediction error from our RL model, orthogonalized with respect to "feedback". We also included temporal derivatives for all of these regressors, six motion parameters (three rotation and three translation), framewise displacement (Power et al., 2014), and six anatomical principal components (Behzadi et al., 2007).

We conducted an analysis in which we looked for neural differences between our conditions within a priori regions of interest (ROI's): the ventral striatum (VS) and ventromedial prefrontal cortex (VMPFC). Both of these ROI's were obtained from the meta-analytic tool Neurosynth (Yarkoni et al., 2011) using the search terms "ventral striatum" and "ventromedial prefrontal cortex". For this analysis we constructed a GLM similar to that described above except delineated by responder group (e.g. opponent $e = 0$, opponent $e = 10$, opponent $e = 20$, expected value $e = 0$, expected value $e = 10$, expected value $e = 20$). We then averaged over runs within subjects, and then took the average BOLD activation within each ROI for each subject. Finally we submitted each subject's averaged BOLD response within the given ROI to a multilevel regression using the lme4 package in R (Bates, Mächler, Bolker, & Walker, 2015; The R Development Core Team, 2017), with BOLD activation within the specific ROI as the dependent variable, social consequence, responder group, and their interaction as independent variables, with the intercept of the model allowed to vary randomly across participants. Significant effects (observed for the VMPFC but not the VS ROI) were followed up with paired *t*-tests and reported in the main text.

Multivariate analysis

We employed multi-voxel pattern analysis (MVPA) to examine patterns of neural activity during both the decision and feedback time-phases in order to detect subtle differences in neural processing between the social and non-social conditions (see **Fig 1A-B**). For each subject we fit a general linear model (GLM) to each trial time-locked to the time-phase of interest (i.e. for both decision and feedback time-phases). This resulted in a single parameter estimate for each trial, during each epoch of interest. We then, for each subject, concatenated these parameter estimates together to create a single image file with 288 volumes, each volume corresponding to the parameter estimate of a given trial. We then applied a 27-voxel searchlight procedure with a linear discriminant analysis (LDA) classifier. A searchlight acts as a traveling region of interest used to detect spatially contiguous patterns of activation specific to functional neural structures (Kriegeskorte, Goebel, & Bandettini, 2006). Each subject completed two social and two non-social functional runs, and we therefore conducted a leave-two-runs-out cross-validation procedure, in which our LDA classifier was trained on two runs (one social and one non-social), and then tested on the two independent left-out runs.

This resulted in a single accuracy map for each subject. The resulting accuracy maps were then concatenated together, and tested for significance with Monte Carlo bootstrapping with 10,000 permutations with threshold-free cluster enhancement (TFCE) (Smith & Nichols, 2009), family-wise error correct (FWE-corrected) at the whole-brain level, implemented in the CoSMoMVPA MATLAB package (Oosterhof, Connolly, & Haxby, 2016). To maximize statistical sensitivity, we created null datasets for significance testing by permuting condition labels. Specifically, for each subject we permuted the labels indicating which condition each volume of their image file belonged to, and ran the searchlight on said permuted image. This process was repeated 100 times for each subject as recommended by Stelzer and colleafgues (Stelzer, Chen, & Turner, 2013) resulting in 100 "null" accuracy maps, representing results from randomized data. These data were used as the null data in the group level TFCE analysis.

Main computational model

Details on BPL

In what follows, we provide mathematical details regarding the derivation of our "Bayesian Preference Learner" or BPL model (see (Devaine & Daunizeau, 2017) for a similar approach). This model essentially describes how Bayesian proposers update, on a trial-by-trial basis, their estimate of the receivers' *acceptance* function parameters. Bayesian proposers assume that the receiver's choices obey a softmax decision rule, which transforms a linear utility function of offer into a probability of accepting the offer (**Fig 8**).

let

$O \in [0{:}20]$ is a potential offer

$s: x \rightarrow 1/(1+\exp(-x))$ be the sigmoid function, (1)

$a_t \in \{0,1\}$ be the receiver's binary choice at trial t,

$f(\theta, O) = \theta_0 + \theta_1 O$ be the estimated (linear) utility that receivers derive from an offer O.

$\theta = \{\theta_0, \theta_1\}$ gathers both the receiver's intercept θ_0 and slope θ_1 of her *acceptance* function.

$s(f(\theta, O)) \triangleq p(a_t = 1|\theta, O)$ is the estimated probability that the receiver accepts an offer O.

Bayesian Learning

Before having observed any receiver's decision, the proposer is endowed with some prior belief $p(\theta)$ about the receiver's behavioral trait θ.

Without loss of generality, we assume that this prior belief $p(\theta) = N(\mu_0, \Sigma_0)$ is Gaussian with mean μ_0 (which captures the direction of the proposer's bias) and variance/covariance Σ_0 (which measures how uncertain is the proposer's prior belief; see **Fig 1, 8**).

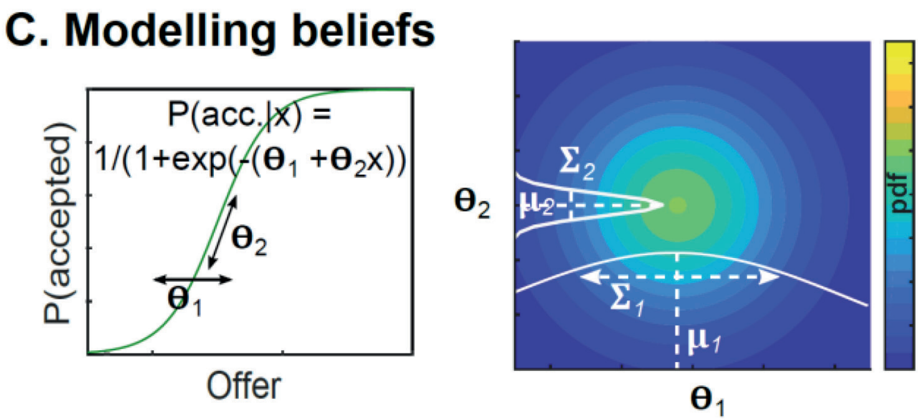

Figure 8. Modeling Beliefs. Each offer has a corresponding probability of acceptance governed by two parameters (θ_1 and θ_2; left panel). Each of these parameters follows a Gaussian distribution described by μ_t and variance Σ_t, which combined create a probability density function (right panel).

Observing the receiver's choices gives the agent information about θ, which can be updated trial after trial using the following Bayes-optimal probabilistic scheme:

$$p(\theta|a_{\rightarrow t}) \propto p(a_{\rightarrow t}|\theta)p(\theta) \tag{2}$$

where $p(\theta|a_{\rightarrow t})$ is the proposer's posterior belief about the receiver's behavioural trait after trial t.
A2 can be rewritten to highlight the trial-by-trial, sequential (online) form of Bayesian learning as:

$$p(\theta|a_{\rightarrow t}) \propto p(a_t|\theta)p(\theta|a_{\rightarrow t-1}) \tag{3}$$

In other words, after observing receiver's decision a_t, the proposer can update her (posterior) belief about the receiver's behavioral trait $p(\theta|a_{\rightarrow t})$, by combining the likelihood of observing the decision given her preceding belief about the receiver's behavioral trait $p(a_t|\theta)$ with her preceding belief about the receiver's behavioral trait $p(\theta|a_{\rightarrow t-1})$.

Equation 3 can be approximated using a variational-Laplace scheme, which essentially replaces the integration implicit in Equation 3 with an optimization of the sufficient statistics of the approximate posterior distributions (Daunizeau, Adam, & Rigoux, 2014; Friston, Mattout, Trujillo-Barreto, Ashburner, & Penny, 2007). This eventually yields semi-analytical expressions for the trial-by-trial update rules of two first moments of the posterior probability density function. In brief, we approximate the posterior belief $p(\theta|a_{\rightarrow t}) \approx N(\mu_t, \Sigma_t)$ in terms of a Gaussian distribution with mean μ_t and variance Σ_t (**Fig. 8**).

Given the observed decision a_t to the offer O_t made at trial t, this leads to the following learning (update) rules for the belief about the receiver's behavioral trait:

$$\Sigma_t = \left(\Sigma_{t-1}^{-1} + s(f(\mu_{t-1}, O_t))\left(1 - s(f(\mu_{t-1}, O_t))\right)\nabla f|_{\mu_{t-1}}{}^{T} \nabla f|_{\mu_{t-1}}\right)^{-1}$$
$$\mu_t = \mu_{t-1} + \Sigma_t \nabla f|_{\mu_{t-1}}\left(a_t - s(f(\mu_{t-1}, O_t))\right) \qquad (4)$$

Critically, and paralleling simpler models in reinforcement learning, it can be seen from Equation A4 that the change in the agent's posterior mean $\mu_t - \mu_{t-1}$ is driven by a *choice prediction error* $(a_t - s(f(\mu_{t-1}, O_t)))$, whose impact is modulated by the agent's subjective uncertainty Σ_t.

Also, note that the proposer's posterior uncertainty about the receiver's behavioral trait Σ_t is monotonically decreasing over trials.

Iterated through time or trials, Equation A4 essentially describes how the proposer learns about the receiver's probability to accept any offer (**Fig. 9**). We refer the interested reader to (Devaine & Daunizeau, 2017; Devaine, Hollard, & Daunizeau, 2014; Mathys, Daunizeau, Friston, & Stephan, 2011) for further mathematical details regarding the derivations of similar meta-Bayesian learning rules.

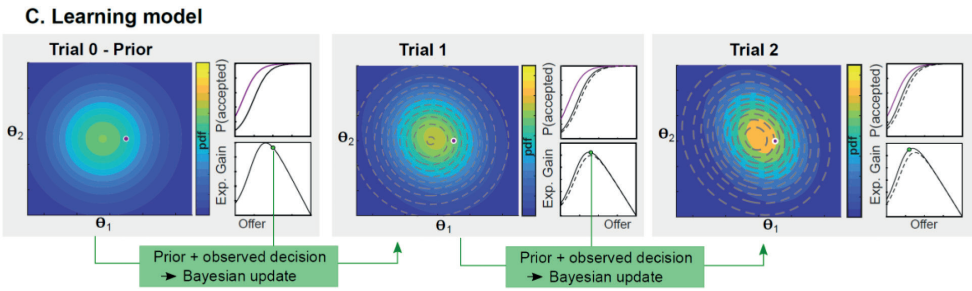

Figure 9. Learning model. Over trials subjects use Bayesian updating to refine their estimates of the parameters governing the acceptance function of the responder group. Consider a proposer, represented by her belief probability density function (colored surface), confronted with a specific culture, represented by the parameters of its acceptance function (purple dot). At each trial, the proposer uses her estimated acceptance function (black curve; top right insets) to produce an expected gain function (bottom right insets), and select an offer that (soft) maximize expected gain (green dot). In this case, the offers are accepted, and the proposer use this information to update her beliefs using Bayes rule. Note that the peak of the belief probability density function gets closer to the true parameters (purple dots). Dotted lines represent the previous trial features.

To summarize, the learning module of our computational model approximate Bayesian optimal learning of receivers' parameters via a variational-Laplace scheme. The free-parameters of the learning module (that can be adjusted/fitted to account for our participants behavior) are the prior beliefs about receiver's intercept θ_0 and slope θ_1 of her *acceptance* function:

$\mu_0 = [\mu_{1,0}, \mu_{2,0}]$; where $\mu_{1,0}$ is the mean of the prior beliefs about the receiver's intercept and $\mu_{2,0}$ is the mean of the prior beliefs about the receiver's slope.

$\Sigma_0 = \begin{bmatrix} \Sigma_{1,0} & \Sigma_{12,0} \\ \Sigma_{12,0} & \Sigma_{2,0} \end{bmatrix}$; where $\Sigma_{1,0}$ is the variance of the prior beliefs about the receiver's intercept and $\Sigma_{2,0}$ is the mean of the prior beliefs about the receiver's slope. $\Sigma_{12,0}$ is the covariance between the prior intercept and slope, which we assume is 0. Hence:

$\Sigma_0 = \begin{bmatrix} \Sigma_{1,0} & 0 \\ 0 & \Sigma_{2,0} \end{bmatrix}$;

Note however that due to learning, the covariance between the prior intercept and slope $\Sigma_{12,0}$ becomes non-zero after the first feedback.

Expected gain

Given the receiver's choices up to trial *t*, the proposer can now form a prediction about the Other's probability to accept any offer O at trial *t* +1:

$$E[a_{t+1}|O_{t+1}, a_{\to t}] = E\left[s\big(f(\theta, O_{t+1})\big)\middle|a_{\to t}\right] \qquad (5)$$

For simplicity we will assume that the predictions only depend on the parameter estimated mean μ_t (note, however, that the variance of the parameters could be included in these predictions).

$$E[a_{t+1}|O_{t+1}, a_{\to t}] = s(f(\mu_t, O_{t+1})) \qquad (6)$$

These predictions can be used to evaluate offer's expected payoff, given the Ultimatum Game payoff matrix:

$$EG[O_{t+1}, a_{\to t}] = (e - O_{t+1}) \times s(f(\mu_t, O_{t+1})) \qquad (7)$$

Performing this evaluation across the whole offer space, receivers can identify which offer O_{t+1} yield the highest expected gain (**Fig. 10**)

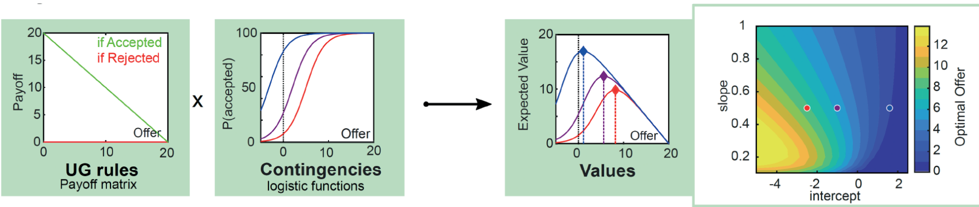

Figure 10. Maximizing expected gain. The ultimatum payoff structure (left panel) is combined with acceptance functions (second panel from left) to derive expected gain, as a function of offers, for each responder group (third panel from left). The optimal policy is to select the offer with the maximum expected gain (diamonds). These offers correspond to specific combinations of estimates of the responder group's slope and intercept (rightmost panel).

Social preferences: inequality aversion

In the social condition participants are not only attempting to maximize their monetary gain, but also taking into account the monetary situation of their opponent. Formally this can be represented by introducing an inequity aversion term into the calculation of utility (Fehr & Schmidt, 1999). Inequity aversion simply reduces the value of outcomes that benefit one party more than the other, with the term itself simplified as follows:

$$ineq = -((O - 10)^2/100) \tag{8}$$

Which means that the payoff from every offer is reduced by a percentage centered around a "fair" 50/50 split. This is incorporated into the calculation of expected gain as follows:

$$EG[O_{t+1}, a_{\rightarrow t}] = (e - O_{t+1}) \times s\big(f(\mu_t, O_{t+1})\big) + ineq \times A \tag{9}$$

Where A is the weight given to the inequity aversion term (see **Fig. 2**).

Author Note and Acknowledgements

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (AdG agreement n° 785635) to CKWDD, a Netherlands Science Foundation VENI Grant (016. Veni.195.078) to JG, and a Swiss National Fund Ambizione grant (PZ00P3_174127) to ML. MRG, ML, JG and CKWDD conceived of the project, designed the studies, and wrote the paper. MRG and AF conducted experiments, and ML, JD, and MRG performed computational modeling and analyses. The authors declare no conflict of interest.

References

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., … Varoquaux, G. (2014). Machine learning for neuroimaging with scikit-learn. *Frontiers in Neuroinformatics*. https://doi.org/10.3389/fninf.2014.00014

Adair, W. L., Okumura, T., & Brett, J. M. (2001). Negotiation behavior when cultures collide: The United States and Japan. *Journal of Applied Psychology*. https://doi.org/10.1037/0021-9010.86.3.371

Avants, B. B., Epstein, C. L., Grossman, M., & Gee, J. C. (2008). Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*. https://doi.org/10.1016/j.media.2007.06.004

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Baumgartner, T., Fischbacher, U., Feierabend, A., Lutz, K., & Fehr, E. (2009). The Neural Circuitry of a Broken Promise. *Neuron*. https://doi.org/10.1016/j.neuron.2009.11.017

Behzadi, Y., Restom, K., Liau, J., & Liu, T. T. (2007). A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2007.04.042

Bicchieri, C. (2005). *The grammar of society: The nature and dynamics of social norms. The Grammar of Society: The Nature and Dynamics of Social Norms*. https://doi.org/10.1017/CBO9780511616037

Blake, P. R., McAuliffe, K., Corbit, J., Callaghan, T. C., Barry, O., Bowie, A., … Warneken, F. (2015). The ontogeny of fairness in seven societies. *Nature*, *528*(7581), 258–261. https://doi.org/10.1038/nature15703

Buchan, N. R., Croson, R. T. A., & Johnson, E. J. (2004). When do fair beliefs influence bargaining behavior? Experimental bargaining in Japan and the United States. *Journal of Consumer Research*. https://doi.org/10.1086/383433

Chen, D. L., Schonger, M., & Wickens, C. (2016). oTree-An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*. https://doi.org/10.1016/j.jbef.2015.12.001

Crawford, V. P. (2019). Experiments on Cognition, Communication, Coordination, and Cooperation in Relationships. *Annual Review of Economics*. https://doi.org/10.1146/annurev-economics-080218-025730

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis: I. Segmentation and surface reconstruction. *NeuroImage*. https://doi.org/10.1006/nimg.1998.0395

Daunizeau, J., Adam, V., & Rigoux, L. (2014). VBA: A Probabilistic Treatment of

Nonlinear Models for Neurobiological and Behavioural Data. *PLoS Computational Biology*. https://doi.org/10.1371/journal.pcbi.1003441
Daw, N. D. (2011). Trial-by-trial data analysis using computational models. *Decision Making, Affect, and Learning: Attention and Performance XXIII*, 1–26. https://doi.org/10.1093/acprof:oso/9780199600434.003.0001
De Dreu, C. K. W., Gross, J., Fariña, A., & Ma, Y. (2020). Group Cooperation, Carrying-Capacity Stress, and Intergroup Conflict. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2020.06.005
De Dreu, C. K. W., Weingart, L. R., & Kwon, S. (2000). Influence of social motives on integrative negotiation: A meta-analytic review and test of two theories. *Journal of Personality and Social Psychology*. https://doi.org/10.1037/0022-3514.78.5.889
Debove, S., Baumard, N., & André, J. B. (2016). Models of the evolution of fairness in the ultimatum game: A review and classification. *Evolution and Human Behavior*. https://doi.org/10.1016/j.evolhumbehav.2016.01.001
Devaine, M., & Daunizeau, J. (2017). Learning about and from others' prudence, impatience or laziness: The computational bases of attitude alignment. *PLoS Computational Biology*, *13*(3), 1–28. https://doi.org/10.1371/journal.pcbi.1005422
Devaine, M., Hollard, G., & Daunizeau, J. (2014). The Social Bayesian Brain: Does Mentalizing Make a Difference When We Learn? *PLoS Computational Biology*, *10*(12). https://doi.org/10.1371/journal.pcbi.1003992
Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., … Gorgolewski, K. J. (2019). fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nature Methods*. https://doi.org/10.1038/s41592-018-0235-4
Fehr, E., & Fischbacher, U. (2002). Why social preferences matter - The impact of non-selfish motives on competition, cooperation and incentives. *Economic Journal*. https://doi.org/10.1111/1468-0297.00027
Fehr, E., & Schmidt, K. M. (1999). A Theory of Fairness, Competition, and Cooperation. *The Quarterly Journal of Economics*, *114*(3), 817–868. https://doi.org/10.1162/003355399556151
FeldmanHall, O., & Shenhav, A. (2019). Resolving uncertainty in a social world. *Nature Human Behaviour*. https://doi.org/10.1038/s41562-019-0590-x
Fiske, A. P. (1992). Fiske-Four Elementary Forms of Sociality 1992.pdf. *Psychological Review*.
Fonov, V., Evans, A., McKinstry, R., Almli, C., & Collins, D. (2009). Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *NeuroImage*. https://doi.org/10.1016/s1053-8119(09)70884-5
Friston, K., Mattout, J., Trujillo-Barreto, N., Ashburner, J., & Penny, W. (2007). Variational free energy and the Laplace approximation. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2006.08.035

Gelfand, M. J., Erez, M., & Aycan, Z. (2007). Cross-cultural organizational behavior. *Annual Review of Psychology*. https://doi.org/10.1146/annurev.psych.58.110405.085559

Gelfand, M. J., & Harrington, J. R. (2015). The Motivational Force of Descriptive Norms: For Whom and When Are Descriptive Norms Most Predictive of Behavior? *Journal of Cross-Cultural Psychology*. https://doi.org/10.1177/0022022115600796

Gorgolewski, K., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., & Ghosh, S. S. (2011). Nipype: A flexible, lightweight and extensible neuroimaging data processing framework in Python. *Frontiers in Neuroinformatics*. https://doi.org/10.3389/fninf.2011.00013

Greve, D. N., & Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2009.06.060

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

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., … Tracer, D. (2005). "Economic man" in cross-cultural perspective: behavioral experiments in 15 small-scale societies. *The Behavioral and Brain Sciences*, *28*(6), 795–815; discussion 815-55. https://doi.org/10.1017/S0140525X05000142

King-Casas, B., Tomlin, D., Anen, C., Camemer, C., Quartz, S., & Montague, 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

Klein, A., Ghosh, S. S., Bao, F. S., Giard, J., Häme, Y., Stavsky, E., … Keshavan, A. (2017). Mindboggling morphometry of human brains. *PLoS Computational Biology*. https://doi.org/10.1371/journal.pcbi.1005350

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences*, *103*(10), 3863–3868. https://doi.org/10.1073/pnas.0600244103

Liu, Y., Huang, Y., Luo, Y., & Zhao, Y. (2012). How does justice matter in achieving buyer-supplier relationship performance? *Journal of Operations Management*. https://doi.org/10.1016/j.jom.2012.03.003

Mathys, C., Daunizeau, J., Friston, K. J., & Stephan, K. E. (2011). A Bayesian foundation for individual learning under uncertainty. *Frontiers in Human Neuroscience*. https://doi.org/10.3389/fnhum.2011.00039

Oosterbeek, H., Sloof, R., & van de Kuilen, G. (2004). Cultural differences in ultimatum experiments: Evidence from a meta-analysis. *Experimental Economics*, *7*(2), 171–188.

Oosterhof, N. N., Connolly, A. C., & Haxby, J. V. (2016). CoSMoMVPA: Multi-Modal Multivariate Pattern Analysis of Neuroimaging Data in Matlab/GNU Octave.

Frontiers in Neuroinformatics, *10*. https://doi.org/10.3389/fninf.2016.00027

Palminteri, S., Wyart, V., & Koechlin, E. (2017). The Importance of Falsification in Computational Cognitive Modeling. *Trends in Cognitive Sciences*, *21*(6), 425–433. https://doi.org/10.1016/j.tics.2017.03.011

Power, J. D., Mitra, A., Laumann, T. O., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2014). Methods to detect, characterize, and remove motion artifact in resting state fMRI. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2013.08.048

Rai, T. S., & Fiske, A. P. (2011). Moral Psychology Is Relationship Regulation: Moral Motives for Unity, Hierarchy, Equality, and Proportionality. *Psychological Review*. https://doi.org/10.1037/a0021867

Rand, D. G., & Nowak, M. A. (2013). Human cooperation. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2013.06.003

Sam, D. L., & Berry, J. W. (2010). Acculturation: When individuals and groups of different cultural backgrounds meet. *Perspectives on Psychological Science*. https://doi.org/10.1177/1745691610373075

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.

Slonim, R., & Roth, A. E. (1998). Learning in High Stakes Ultimatum Games: An Experiment in the Slovak Republic. *Econometrica*. https://doi.org/10.2307/2998575

Smith, S. M., & Nichols, T. E. (2009). Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage*, *44*(1), 83–98.

Soares, M. C., Bshary, R., Fusani, L., Goymann, W., Hau, M., Hirschenhauser, K., & Oliveira, R. F. (2010). Hormonal mechanisms of cooperative behaviour. *Philosophical Transactions of the Royal Society B: Biological Sciences*. https://doi.org/10.1098/rstb.2010.0151

Stelzer, J., Chen, Y., & Turner, R. (2013). Statistical inference and multiple testing correction in classification-based multi-voxel pattern analysis (MVPA): Random permutations and cluster size control. *NeuroImage*, *65*, 69–82. https://doi.org/10.1016/j.neuroimage.2012.09.063

The R Development Core Team. (2017). R: A language and environment for statistical computing.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*. https://doi.org/10.1017/S0140525X05000129

Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., & Gee, J. C. (2010). N4ITK: Improved N3 bias correction. *IEEE Transactions on Medical Imaging*. https://doi.org/10.1109/TMI.2010.2046908

Weber, R. A., & Camerer, C. F. (2003). Cultural conflict and merger failure:

An experimental approach. *Management Science*. https://doi.org/10.1287/mnsc.49.4.400.14430

Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., & Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nature Methods*, *8*(8), 665–670. https://doi.org/10.1038/nmeth.1635

Zhang, Y., Brady, M., & Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*. https://doi.org/10.1109/42.906424

A

Appendix

Summary
Samenvatting
Acknowledgements

Summary

This thesis consists of three empirical chapters that investigate elements of human social behavior through the combination of economic games, computational modeling, and neuroimaging. **Chapter 2** uses the attacker-defender contest and a cognitive-hierarchies framework (Camerer, Ho, & Chong, 2004). The cognitive-hierarchies framework quantifies the depth of mentalizing recursion, i.e., I think that you think that I think. We found that during attack relative to defense individuals invested less and were less successful, and that investments in attack utilized more levels of cognitive recursion (i.e. more sophisticated mentalizing) than investments in defense. Furthermore, attack behavior was preferentially associated with neural activity in the ventral striatum, a region consistently linked with reward learning, and the temporoparietal junction, a region consistently linked with perspective-taking and social cognition. We conclude that in economic contests, coming out ahead (versus not falling behind) involves sophisticated strategic reasoning that engages neural regions associated with both value computation and theory of mind.

A key task for defenders in the attacker-defender game studied in Chapter 2 is to assess to what extent they can trust their counterpart to not attack, or should instead fear their counterpart's aggressiveness. **Chapter 3** zooms in on trust and distrust as a key element in social interactions. We show that variability in reciprocity (participants playing as responders) can be exhaustively captured by three categories: exploiters (individuals who never reciprocate), perfect reciprocators (individuals who always reciprocate), and contingent reciprocators (individuals who reciprocate as a function of how much they are trusted). This variability is learned by senders through a combination of reinforcement and belief-based learning. However, senders learn to trust imperfectly, frequently failing to arrive at the optimal policy, in particular when interacting with contingent reciprocators. Furthermore, the degree to which individuals weigh belief over reinforcement is positively correlated to their average payoff, indicating that learning to trust from mentally simulated outcomes outperforms learning from observation only.

The results from both Chapter 2 and 3 revealed an important role for social perception and learning, suggesting that empathy and social norms modulate decisions to exploit and to trust and reciprocate. **Chapter 4** builds on these and related findings by asking what role empathy (Zaki, 2014; Zaki & Mitchell, 2013) and social preferences such as concerns for fairness and the welfare of others (Blake et al., 2015; Fehr & Schmidt, 1999) play in learning group-specific conventions. We created three populations with different rules of engagement and varied whether or not decisions affected interaction partner outcomes. Participants made ultimatum bargaining offers to responders from these different populations and could observe whether their offer was accepted or rejected. Participants quickly adapted to group-specific rules in learning environments without social consequences, but were overly generous and ended up misrepresenting

Appendix

what would be acceptable when decisions affected their partner's outcomes. We propose a computational model, combining Bayesian principles and social preferences, that mechanistically explains how generosity leads to biased sampling, impeded learning, and false beliefs about what offers are deemed acceptable. Using functional neuroimaging, we mapped key computational variables in two major brain networks, previously associated with value-based and social decision-making. Results suggest that generosity, related to brain regions associated with decision-conflict and perspective-taking, can induce self-fulfilling beliefs in pro-sociality norms that may help to increase cooperation and reduce conflict between distinct groups but also create inaccurate stereotypes and economic inefficiencies.

Samenvatting

Dit proefschrift bestaat uit drie empirische hoofdstukken die elementen van menselijk sociaal gedrag onderzoeken door de combinatie van economische spellen, computationele modellen, en neuroimaging. Hoofdstuk 2 maakt gebruik van de attacker-defender contest en een raamwerk van cognitieve hiërarchieën (Camerer, Ho, & Chong, 2004). Het raamwerk van cognitieve hiërarchieën kwantificeert het niveau van mentaliserende recursie, d.w.z., "ik denk dat jij denkt dat ik denk". We ontdekten dat tijdens een aanval individuen minder investeerden en minder succesvol waren dan tijdens een verdediging, en dat investeringen in een aanval meer niveaus van cognitieve recursie (d.w.z., een geavanceerdere mentalisatie) gebruikten dan investeringen in een verdediging. Bovendien was aanvalsgedrag bij voorkeur geassocieerd met neurale activiteit in het ventrale striatum, een regio die consequent verbonden is met beloningsleren, en de temporoparietale junctie, een regio die consequent verbonden is met perspectief nemen en sociale cognitie. We concluderen dat in economische spellen, vooruit blijven (in plaats van niet achterblijven) geavanceerde strategische redeneringen worden gebruikt die neurale regio's stimuleren die verband houden met zowel waardeberekening als theory of mind.

Een belangrijke taak voor verdedigers in de attacker-defender contest die in hoofdstuk 2 wordt bestudeerd, is om te beoordelen of ze erop kunnen vertrouwen dat de tegenpartij niet aanvalt, of dat ze bang moeten zijn voor agressiviteit. Hoofdstuk 3 zoomt in op vertrouwen en wantrouwen als sleutelelement in sociale interacties. We laten zien dat variatie in wederkerigheid (van individuen die als respondenten spelen in het trust-spel) kan worden samengevat in drie categorieën: uitbuiters (individuen die nooit wederkerig zijn), perfecte wederkerigheid (individuen die altijd wederkerig zijn) en contingente wederkerigheid (individuen die wederkerig zijn in functie van hoeveel ze worden vertrouwd). Deze variabiliteit wordt door afzenders in het trustspel aangeleerd door een combinatie van bekrachtiging en leren op basis van overtuigingen. Afzenders leren niet volledig te vertrouwen en komen vaak niet tot de optimale uitkomst, met name wanneer ze met spelers van de categorie contingente wederkerigheid omgaan. Bovendien is de mate waarin de overtuiging van individuen zwaarder weegt dan bekrachtiging positief gecorreleerd met hun gemiddelde beloning, wat aangeeft dat vertrouwen aanleren op basis van mentaal gesimuleerde uitkomsten betere uitkomsten oplevert dan leren door alleen observatie.

De resultaten van zowel hoofdstuk 2 als hoofdstuk 3 lieten een belangrijke rol voor sociale perceptie en leren zien, wat suggereert dat empathie en sociale normen beslissingen moduleren die belangrijk zijn voor exploitatie, vertrouwen, en wederkerigheid. Hoofdstuk 4 bouwt voort op deze en gerelateerde bevindingen door in te gaan op de rol die empathie (Zaki, 2014; Zaki & Mitchell, 2013) en sociale voorkeuren over rechtvaardigheid en het welzijn van anderen (Blake et al., 2015; Fehr &

Schmidt, 1999) spelen in het leren van groepsspecifieke conventies. We creëerden drie populaties met verschillende interactieregels en varieerden of beslissingen al dan niet van invloed waren op de resultaten van interactiepartners. Deelnemers deden aanbiedingen in het ultimatumspel aan respondenten uit deze verschillende populaties en konden observeren of hun aanbod werd aanvaard of afgewezen. Deelnemers pasten zich snel aan wat betreft de groepsspecifieke regels in leeromgevingen zonder sociale gevolgen, maar waren overdreven genereus en gaven uiteindelijk een verkeerde voorstelling van wat acceptabel zou zijn als beslissingen van invloed waren op de resultaten van hun partner. We introduceren een computationeel model dat Bayesiaanse principes en sociale voorkeuren combineert en mechanistisch uitlegt hoe vrijgevigheid leidt tot vertekende steekproeven, beperkt leren, en foute overtuigingen over wat voor aanbiedingen als acceptabel worden beschouwd. Met behulp van neuroimaging hebben we de belangrijkste computationele variabelen in kaart gebracht in twee grote hersennetwerken, die voorheen werden geassocieerd met op waarden gebaseerde en sociale besluitvorming. De resultaten suggereren dat vrijgevigheid, gerelateerd aan hersengebieden die verband houden met beslisconflicten en perspectief nemen, zelfvervullende overtuigingen over pro-socialiteitsnormen kan opwekken die kunnen helpen om samenwerking te vergroten en conflicten tussen verschillende groepen te verminderen, maar ook kunnen leiden tot onnauwkeurige stereotypen en economische inefficiënties.

Acknowledgements

This dissertation would have been impossible without an enormous amount of support. First and foremost I would like to thank my promoter, Carsten De Dreu, who allowed me an incredible amount of freedom in pursuing topics that I found interesting, even when those topics were not within the original plan for my dissertation. Much of my current understanding of psychology, economics, and how the brain reasons in the social world is owed to the many hours we spent in conversation unpacking complex ideas and strategizing on how best to answer fascinating scientific questions. Your confidence in me and patience throughout my diversions down many rabbit holes and up against many expensive road blocks have completely spoiled me as an employee, and have infused me with optimism and enthusiasm as a researcher.

I owe a debt of gratitude to my supervisor, Jörg Gross, whose intimidating intellect was a constant source of motivation. I rarely came to you with a theoretical, methodological, or practical problem that you couldn't solve, but more importantly the casual conversations over drinks oftentimes left me reeling with curiosity and with new ways of thinking—the value of which cannot be overstated.

To my primary collaborator and methodological mentor, Maël Lebreton, who patiently taught me everything I know about the nuts and bolts of computational modeling. Since our first coding session I've made an effort to include you on as many projects as possible to continue mining you for more knowledge. This dissertation would look much different (and worse) without your involvement.

I also could not have done this project without Daniel Lindh, whose relentless overenthusiasm, oftentimes exhausting, nonetheless manacled me to hope throughout many crises of faith. I owe a massive debt of gratitude to Eliska and Luisa Prochazkova for their unflinching attention to detail and for endless discussions on how we can solve all the world's problems with our research projects. And to Renae McCollum, for bringing out my inner optimist even while I sat cursing the world of science.

I owe a huge debt to Hilmar Zech, the statictician who helped me climb the ranks of analytical thinking and helped execute countless beta pipelines. To Andrea Fariña for being at my side through almost all of these projects, never missing a chance gleefully tell me I'm wrong and correct my mistakes, and making it easy to come to work. To Andrea Arciniegas for nudging me to expand my views on nearly every topic, despite my stubbornness. To Camila Miorelli for millions of mind-blowing conversations about the tumultuous love affair between science and art. To Leon Hilbert, Ruthie Pliskin, Lennart Reddmann, Mirre Stallen, Zsombor Méder, Angelo Romano, Maria Lojowska, Laura Hoenig, Thies Gehrmann, Geert-Jan Will, Josipa Alilovic, Sasha Libina, Josi Marshall, Leon Diaz De Santillana, Anne van Stapele, Auke van der Veen, Olmo van den Akker, Anthony Passantino, Dylan and Cole Titus, Elío Sjak-Shie, Jordan Wilson, Tyler Johnson, Tyler Judson, Ryan Tamborski, and the rest of the Leiden University staff.

To Jason Throop for the paternal guidance and intellectual rigor he applied to my ideas. To my sisters Jessica and Rachel for their unwavering support and for being amazing sisters. And to Jon Drappelik and Steve Strong for actually choosing to be part of the chaos.

To Buck and Dotty Reesor for praising me as a genius since I was a child, and Cameron and Cailie Reesor, my little brother and sister, for being actual geniuses and thus allowing me to believe the praise. To Auntie Anne and Uncle Les, Matt and Kim, Uncle John and Aunt Kathy, and everyone else saying Jaja's prayer once a year.

Most importantly, to my parents for giving me the freedom to pursue any goal I want, for truly instilling in me the implication of that freedom, and for not failing to show pride and support at all my accomplishments and for picking me up after all my defeats. If it were even possible to put into words the love and gratitude and appreciation and luckiness I feel about having such perfect parents, I certainly wouldn't have space for it here.