How neurons, norms, and institutions shape group cooperation

Jay J. Van Bavel
New York University

Philip Pärnamets
Karolinska Institutet; New York University

Diego Reinero
Princeton University

Dominic Packer
Lehigh University

Van Bavel, J. J., Pärnamets, P., Reinero, D., & Packer, D. (in press). How neurons, norms, and institutions shape group cooperation. *Advances in Experimental Social Psychology*

Word count: 17,950 (total)

Acknowledgments and author's note: This chapter was partially funded by a grant from the John Templeton Foundation to JVB & PP. PP was also supported by Swedish Research Council grants 2014-01371 and 2020-02584.

Please direct correspondence to:
Jay J. Van Bavel
New York University
Department of Psychology
New York, NY, USA. 10012
jay.vanbavel@nyu.edu

Abstract

Cooperation occurs at all stages of human life and is necessary for small groups and large-scale societies alike to emerge and thrive. This chapter bridges research in the fields of cognitive neuroscience, neuroeconomics, and social psychology to help understand group cooperation. We present a value-based framework for understanding cooperation, integrating neuroeconomic models of decision-making with psychological and situational variables involved in cooperative behavior, particularly in groups. According to our framework, the ventromedial prefrontal cortex serves as a neural integration hub for value computation during cooperative decisions, receiving inputs from various neuro-cognitive processes such as attention, affect, memory, and learning. We describe factors that directly or indirectly shape the value of cooperation decisions, including cultural contexts and social norms, personal and social identity, and intergroup relations. We also highlight the role of economic, social, and cultural institutions in shaping cooperative behavior. We discuss the implications for future research on cooperation.

Keywords: Cooperation, Groups, Identity, Decision-making, Pro-Sociality, Dual-Process, Norms, Intragroup, Intergroup

1. Introduction

Cooperation occurs at all stages of human life and is necessary for small groups and large-scale societies alike to emerge and thrive. Cooperation is observed early in human development and is fundamental to the success of relationships, small teams, large organizations, and nations. When individuals prioritize themselves at the expense of others, the consequences can damage social communities, scientific institutions, and the planet. Understanding why people cooperate with others—or fail to do so—is a central question for the social and biological sciences, as well as policy makers and leaders.

Debates about the nature of human cooperation and prosociality have raged for centuries. For instance, philosophers have debated whether prosocial tendencies are rooted in institutions that regulate our selfish impulses (Hobbes, 1651/2002) or develop more organically as part of human nature (Rousseau, 1754/1998). In recent years, the study of cooperation has evolved beyond philosophical debate and consumed the energy of scientists across numerous disciplines, from primatology to economics, cell biology to international relations. Interest in the dynamics of cooperation is also fundamental to the field of social psychology, which focuses on the role of psychological and situational factors underlying cooperation.

One on hand, models of *prosocial restraint* assert that the better angels of our nature stem from deliberate restraint of deeply rooted self-interest (DeWall, Baumeister, Gailliot, & Maner, 2008; Kocher, Martinsson, Myrseth, & Wollbrant, 2012; Stevens & Hauser, 2004). On the other hand, models of *prosocial intuition* argue that cooperation is often a spontaneous response, which can be corrupted by deliberate attempts to

maximize self-interest (Rand, 2016; Rand, Greene, & Nowak, 2012). This debate has inspired dozens of studies and a mixed body of evidence in the cooperation literature, which eventually led to a large-scale international replication project to try to answer the question–which itself produced mixed results (see below for more details; Bouwmeester et al., 2017). Unfortunately, the extensive research devoted to the mental processes underlying group cooperation has failed to produce a satisfying model for understanding how the human mind decides to cooperate. The current chapter bridges research in the fields of cognitive neuroscience, neuroeconomics, and social psychology to help understand group cooperation.

Central to the issue of cooperation is understanding who can (and cannot) be trusted. There are at least three fundamental “modes of trust production” necessary for cooperation (Zucker, 1986) that we will review in this chapter: (1) the trust and resulting cooperation that develops from direct experience with and reputational knowledge about individuals, (2) the trust that arises due to shared group identities or from otherwise categorizing individuals, and (3) institutional trust, in which “formal mechanisms are used to provide trust that does not rest on personal characteristics or a particular history of exchange” (Zucker, 1986, p. 61). Each of these factors can underlie our willingness to trust others during cooperative decision-making.

We ground our perspective in the neuroeconomic literature on value-based decision-making. The study of decisions based on subjective valuations, from deciding whether one wants coffee or tea with breakfast to choices between cooperation or defection with group members, suggests that they are based on a set of domain general neurocognitive processes (Levy & Glimcher, 2012). These processes are used to

determine the value of competing decision options and implement choices in the real world. We believe this work provides an integrative framework for research across disciplinary boundaries, and offers deeper insights into the diversity of processes and factors that can impact cooperation (Wilson, 1998).

Specifically, we present a unifying value-based framework for understanding cooperation, integrating neuroeconomic models of decision-making with psychological and situational variables involved in cooperative behavior, particularly in groups (see Pärnamets et al., 2020).[1] According to our framework, the ventromedial prefrontal cortex serves as a neural integration hub for value computation during a wide variety of decisions (including cooperation), receiving inputs from various neuro-cognitive processes such as attention, affect, memory, and learning. Our framework also incorporates factors that directly or indirectly shape the value of cooperation decisions, including cultural contexts and social norms, personal and social identity, and intergroup relations. We also highlight the role of economic, social, and cultural institutions in shaping cooperative behavior. In doing so, we help advance theoretical debates about cooperation by explaining how previous, often contradictory, findings can be accommodated within a general value-based framework. Finally, we highlight domains for future research on cooperation in social psychology.

In the first section, we describe the theoretical value-based framework underlying human cooperation. This section will revisit evidence for dual-process models in light of value-based decision-making and introduce the role of social cognition in cooperative decision-making at multiple levels of analysis. We then discuss the role of social identity

[1] The current chapter focuses primarily on group and large-scale cooperation, as opposed to dyadic forms of cooperation (which often rely on different tasks and experimental paradigms),

and norms on cooperation. Next, we examine the impact of macro-level social factors, such as cultural and institutional factors. In the last section we discuss the future of cooperation. This includes the role of inequality, democracy, and climate change as contexts that influence cooperation as well as problems that will require wide scale cooperation.

2. Value-based decision-making

When making decisions, from choosing what snack to purchase at a vending machine to deciding whether to donate to charity, the brain must assign or construct values for all available options in accordance with an individual's preferences. For values of different options to be comparable they must be translated to a domain-general neural code or "common currency". Using such subjective values allows people to compare options that may differ in any number of features, trade-off different values (e.g., self-interest vs. social value), and engage in goal-directed behavior. A large body of work in neuroscience has revealed a core valuation circuit, consisting of the orbitofrontal cortex, ventromedial prefrontal cortex (vmPFC), and ventral striatum, in which value representation and computation occurs (Bartra, McGuire, & Kable, 2013; Levy & Glimcher, 2012). These neuro-cognitive computations are then used to guide decision-making across domains

In a value-based framework, the values of actions when facing cooperative dilemmas are assumed to be constructed based on multiple inputs to the core valuation systems. As such, decisions to cooperate engage the valuation system in a similar way to other value-based decisions. For example, when choosing to donate money to strangers, act fairly, or to cooperate, studies have found vmPFC activity correlated with

these decisions (Cutler & Campbell-Meiklejohn, 2019; Fehr & Camerer, 2007). Additionally, brain damage or impairment to the vmPFC can alter prosocial decision-making and change the likelihood for individuals to cooperate (Krajbich, Adolphs, Tranel, Denburg, & Camerer, 2009), such as making it more likely people will cooperate (Wills et al., 2018). As such, the domain general computations used to guide economic decisions also appear to underlie social decisions to engage in selfish or selfless behavior within groups. Thus, one way to understand cooperation is to approach it from the lens of theories and findings relating to value-based decision-making.

The subjective value of different decision options is constructed using inputs from multiple mental processes supported by a diversity of brain regions, including areas associated with basic and higher-order cognitive processes and social functions (see **Figure 1**). Thus, our framework attends to the fact that a wide variety of mental processes can feed into computations of subjective value. Some of these are relatively basic cognitive processes, including perception and attention. Others include higher-order cognitive processes including learning and memory. Still others are social cognitive, including processes involved in representing other minds, tracking social norms, and forming social identities.

It is heuristically useful to think of value computations as involving three distinct parameters. Different choices are valued based on their *potential outcomes* weighted by *estimates of their probability* and decision-makers' *social preferences* (i.e., how concerned they are for their own and others' outcomes; for examples of utility functions see Fehr & Schmidt, 1999; Fehr & Schurtenberger, 2018; Charness & Rabin, 2002). We posit that a variety of basic, higher-order, and social cognitive processes feed

information regarding these parameters into value computations and thereby influence cooperative decisions. The ability to track reputational information about interaction partners is, for example, supported by episodic memories of their prior behavior. Recollections of how people behaved in the past affect expectations for their behavior in the future, thus influencing estimates of the probabilities of different outcomes that are contingent on what those others choose to do. Indeed, one of the key advantages of a value-based framework is that it expands researchers' attention to the role that a wider-range of mental processes can play in cooperation, some of which can provide alternate explanations for findings that are generally interpreted, in psychology, through a dual-process lens. As such, the value-based framework we introduce here can guide integrative research across disciplinary boundaries within psychology by refocusing researchers' attention towards questions of how psychological processes support decision-making in cooperative dilemmas in a way that is consilient with their role in human cognition and social interaction writ large.

When making cooperation decisions, potential *outcomes* to the self and others are weighted by estimates of their *probability*, as well as the degree to which decision-makers are concerned with their own and others' interests (i.e., *social preferences*). These three parameters–*outcomes, probabilities, and social preferences*–provide a useful organizing framework when we consider how both psychological factors (e.g., prosocial tendencies, social identities) and environmental or contextual factors (e.g., social norms, societal institutions) affect cooperative decisions. We briefly illustrate how different factors might influence cooperation through these parameters.

Social norms, for example, may influence individuals' cooperation decisions through all three parameters (Bicchieri, 2005a; Fehr & Schurtenberger, 2018). Because violations of social norms are often sanctioned by communities, they can alter potential outcomes by imposing possible costs on non-normative choices (whether the norms be cooperative or selfish). Social norms may also directly affect individuals' preferences, such that their beliefs about what they should be striving toward in particular situations are influenced by normative patterns of behavior. For example, cooperative social norms might not simply increase cooperation because people expect greater costs for non-cooperation, but because they change what people actually want–causing them to actually place greater value on other people's outcomes (Hackel, Wills, & Van Bavel, 2020). Finally, social norms affect decision-makers' estimates of the probability of different outcomes by providing information about how others are likely to behave. Strong or tight social norms mean that one can expect others to behave in quite predictable ways (Gelfand, 2019). In contrast, in looser normative environments there may be significantly more uncertainty about interaction partners' likely behaviors and thus about the probabilities of different outcomes.

Shared social identities play a particularly central role in shaping cooperation (Kramer & Brewer, 1984). Again, this may occur by affecting more than one parameter. There is substantial evidence that shared identities alter people's social preferences, causing them to extend concern about outcomes from their individual selves to other members of their in-groups (Van Bavel & Packer, 2021). As a result, people tend to be more cooperative with in-group than out-group members across a wide variety of domains. However, this heightened cooperation likely occurs not only because people

care about fellow group members' outcomes, but because they also expect other in-group members to care about theirs (Yamagishi & Kiyonari, 2000).

Understanding that others share our identity is a fundamental element of cooperative decision-making. For instance, one line of research had people play games in which they could choose to trust an in-group or an out-group partner in a money allocation task (Platow, Foddy, Yamagishi, Lim, & Chow, 2011; see also Foddy, Platow & Yamagishi, 2009). In experiments using both real-world and minimal groups, they found that people trusted an in-group member more than an out-group member only when they *believed* that their partner was aware of their group membership. When the in-group partner knew of their shared group membership and could thus be expected to behave prosocially toward them, people were more likely to place their faith in them. When their partner was unaware of the shared identity and would not, therefore, be inclined to care about their interests, they were not more trusting of this person. These mental calculations about shared identity underlie many decisions to cooperate.

In this chapter, we also consider how social institutions like legal systems, regulatory agencies, insurance companies, and other social structures influence cooperation. As with other environmental or contextual influences, institutions can affect more than one parameter in decision-makers' value computations. Institutions can change the set of possible outcomes by developing mechanisms for rewarding or punishing particular behavior (e.g., Gürerk, Irlenbusch & Rockenbach, 2006; Zucker, 1986; Ostrom, 1990). For example, when law-makers outlaw a previously permissible behavior or legalize a formerly illegal action they alter the set of possible outcomes associated with different choices. Institutions can also affect probability estimates by, for

example, altering expectations about how other people are likely to behave (e.g, Zucker, 1986). If there are legal mechanisms for detecting and punishing fraud one may conclude that interaction partners' are less likely to cheat because they are disincentivized from cheating. Finally, certain kinds of institutions might also seek to change people's social preferences. Educational curricula, both religious and secular, are often intended to shape people's moral and ethical beliefs–including the sorts of concerns they should be taking into account when making decisions (Declerck & Boone, 2018). In this way, institutions too can influence cooperation through a number of different pathways.

3. How psychologists study cooperation

Although cooperation spans a broad range of potential activities, from working as a sports team to conducting a study with other scientists, researchers tend to study group cooperation in laboratory settings using stylized interactions known as economic games (cf. Dawes, 1980; Fehr & Camerer, 2007; Curry, 2016). The most commonly used of these economic games includes the Prisoner's Dilemma (PD), the Dictator Game (DG), Ultimatum Game (UG), and the Public Goods Game (PGG), although there are many other games used to study social decision-making (e.g., the Trust Game or Stag Hunt Game; see for example van Dijk & De Dreu, 2021; Thielmann, Böhm, Ott, & Hilbig, 2021). Through their formal structure, economic games help researchers isolate features of social interactions, including possible motivations for prosocial or cooperative decision-making. They also allow scholars to study these decisions in contexts with incentivized choice and real stakes.

A simple cooperative dilemma is represented in the well-known Prisoner's Dilemma. In the Prisoner's Dilemma, two players are faced with two choices – cooperate or defect. If both players cooperate, they both receive a prize (often a few dollars). However, if one player chooses to defect when the other cooperates, the defecting player receives a larger prize, and the cooperating player a loss (a version with costs instead of rewards designed to mimic the payoff structure of prisoners deciding whether to stick together or rat each other out). If both players defect, they both receive a small prize. This payoff structure ensures that players are faced with a temptation to defect, but both players will collectively receive the most rewards if they both decide to cooperate. This tension between the collective good (cooperating) and individual good (defecting) is the hallmark of social dilemmas (Dawes, 1980; Fehr & Camerer, 2007; Van Lange, Joireman, Parks, & van Dijk, 2013) and it also sits at the core of many political debates.

To study cooperation with more than two players, a situation more representative of many important real-life social dilemmas, researchers use the public goods game (PGG; see **Figure 2**). In the PGG, players are endowed with money and are given the choice of contributing some or all of it to a public pool (cooperating) or keeping most or all of it for themselves (defecting). Money given to the public pool is multiplied by the experimenter (e.g., doubled or tripled). Then, the pool is evenly redistributed to all players regardless of their level of cooperation. These features ensure that any individual can maximize their own payoff by withholding contributions to the public good (i.e., defecting) whereas collectively, the most beneficial outcome occurs when everyone cooperates maximally. In a repeated trials version of the PGG,

absent enforcing mechanisms such as social norms, gossip, or punishment, contributions tend to diminish due to diffusion of responsibility or the absence of reciprocity (Andreoni, 1988). This makes the PGG a useful framework for studying how groups can achieve cooperation (as well as what motivates any individual player to cooperate).

While the PGG is similar to the Prisoner's Dilemma, it also introduces additional properties from group psychology given its inherent multiplayer structure (Dawes, 1980). For example, communicative norms such as promise keeping (Bicchieri, 2002), as well as factors like the social identities of the players (Kramer & Brewer, 1984) shape decisions in the PGG.

Cooperating in social dilemmas, like the PGG, requires weighing self-regarding motivations with other-regarding motivations (i.e., concerns for other people's welfare, intentions, and behavior). This tendency is often modeled through the concept of social preferences, we introduced above, which can be understood as parts of a player's utility function that are dependent on the payoff or welfare of others. For instance, some people have a prosocial orientation and care deeply about the welfare of others. Research on cooperation suggests that social preferences can evolve and provide benefits to groups who value the welfare of non-kin strangers (Bowles & Gintis, 2011). Social preferences have been modeled in multiple ways, including aversion towards inequity (Fehr & Schmidt, 1999), guilt (Charness & Rabin, 2002), and norms of conditional cooperation (Bicchieri, 2005b; Fehr & Schurtenberger, 2018). These social preferences inform whether people decide to cooperate (or not) in social dilemmas.

4. Dual-process vs. value-based models of cooperation

One of the longstanding questions in psychology about the nature of cooperation is whether it relies primarily on intuitive or deliberative processes. This dual-process approach to understanding cooperation rests on the assumption that differences between cooperative and selfish decisions reflect the operation of distinct cognitive processes or systems. According to this framework, studies showing that decisions to cooperate are faster compared to decisions to be selfish tend to be interpreted as showing that cooperation is produced by a fast or intuitive cognitive system, relying on heuristics that promote intuitive cooperation (Rand et al., 2012). The alternative dual-process view states the opposite: that people are intuitively selfish and must override these selfish impulses with deliberative thinking (DeWall et al., 2008; Kocher et al., 2012; Martinsson, Myrseth & Wollbrant, 2012; Stevens & Hauser, 2004). The conflict between these two views is one of what is the default approach that humans tend to take in social dilemmas, i.e., is it one of cooperation or of defection? We will argue that both views err in this assumption of there being a definitive default approach; nevertheless, important insights have been generated from the dual-process research, particularly that of the importance of control processes modulating value computations.

Some scholars have proposed a *social heuristics hypothesis* (Rand et al., 2014; Rand, 2016) which makes three claims: (i) cooperating with others is typically advantageous in daily life, and as such, is the default response for most people, (ii) people over-generalize this cooperative “rule of thumb” and default to being cooperative even in atypical contexts when cooperation is seemingly disadvantageous, such as in one-shot anonymous public goods games in the lab, (iii) when people instead intentionally deliberate and reflect on their choices during such atypical contexts, they

realize that there is no incentive for being cooperative and instead become more selfish. Consistent with these hypotheses, initial research on social heuristics found that people cooperate more when asked to make their decisions quickly than when forced to delay their decisions (Rand et al., 2012; Rand et al., 2014). This was taken as evidence that people are *intuitively cooperative*.

However, several other labs had trouble replicating these findings (e.g., Tinghög et al., 2013; Verkoeijen & Bouwmeester, 2014). These mixed results inspired a large-scale and pre-registered replication attempt that combined 21 independent samples from around the world. Each sample relied upon the same basic research design that was agreed to in advance by the original authors. This registered replication attempt did not replicate the intuitive cooperation effect when including all participants (Bouwmeester et al., 2017). However, they were able to replicate the intuitive cooperation effect when excluding what turned out to be the vast majority of participants, who did not comply with the task instructions (see **Figure 3**). This was most often people who did not respond quickly enough in the “time-pressure” condition. Only 34% of participants assigned to the “time-pressure” condition actually complied with this instruction. These excluded participants tended to behave in the exact opposite manner of the social heuristics hypothesis. Taken together, the results pointed toward the possibility that selective exclusion of participants, and not the manipulation of intuition vs. deliberation, may explain the overall intuitive cooperation effect. A recent meta-analysis further supports these conclusions (Kvarven et al., 2020).

Another challenge to dual-process models has emerged from an analysis of reaction times during decision-making. The original work on intuitive cooperation found

that longer reaction times were associated with lower rates of cooperation (Rand et al., 2012). This was initially taken as further evidence that intuition supports cooperation. However, recent work found that patterns of response time data can also be explained by a value-based model of cooperation. Consistent with the value-based approach, response times are related to the discriminability of options (Krajbich, Armel & Rangel, 2010; Shadlen & Kiani, 2013). Discriminability in the case of perceptual decisions (where it is often studied) refers to how perceptually similar two options are. In the case of value-based decisions, discriminability instead relates to how similarly valued the options are by the decision-maker. According to this account, the more similar (i.e. less discriminable) two options are, the longer it will take to choose between them (Evans, Dillon, & Rand, 2015). In other words, the choice is simply more difficult when the options are similar.

There is a large body of evidence that this relationship between response times and option discriminability holds for different types of decisions (e.g., Krajbich, Armel & Rangel, 2010), including social or moral choices (Krajbich, Hare, Bartling, Morishima, & Fehr, 2015; Pärnamets, Balkenius, & Richardson, 2014). A value-based framework appears to provide a more accurate explanation for reaction time differences in cooperation (Krajbich, Bartling, Hare, & Fehr, 2015; Mischkowski & Glöckner, 2016; Teoh, Yao, Cunningham & Hutcherson, 2020). For example, in one experiment, people were invited to play multiple public goods games in a group with other players (Krajbich, Bartling, Hare, & Fehr, 2015). The different games varied the incentive to cooperate, such that the expected return for each unit of money contributed to the public good differed across games. The incentive varied from being very low, and therefore favoring

selfishness, to being high, and therefore favoring cooperation. Past studies on the intuitive cooperation hypothesis (e.g., Rand et al., 2012) used a moderate incentive and found that people who gave more to the group decided faster than those who decided more slowly. However, when cooperation was made relatively costly, the correlation between response times and amount contributed to the public good was positive (meaning faster responders tended to contribute *less*). Conversely, when cooperation was more lucrative, the correlation was instead negative (meaning faster responders tended to contribute more). The ease of a choice (when cooperating is lucrative it is an easy choice) tracks cooperative decisions and explains the association between response times and cooperation. This pattern of results is more consistent with a value-based model of cooperation, where the value of cooperating (versus being selfish) determines the time to make decisions.

The same logic applies to individual differences in value–known as social preferences. Thinking fast boosts cooperation among those who tend to act prosocially, but encourages greed among those who tend to act selfishly. Indeed, prosocial individuals (i.e., those who generally help others) are faster to cooperate than free-ride (i.e., prioritize themselves over the group), whereas selfish individuals are faster to free-ride than cooperate (Hutcherson, Bushong, & Rangel, 2015; Krajbich, Bartling, Hare, & Fehr, 2015). For instance, researchers measured how prosocial (vs. selfish) participants were using an index called Social Value Orientation (i.e., a person's preference about how to allocate resources between the self and another person; see Van Lange, De Bruin, Otten, & Joireman, 1997). Participants were then asked to make contributions to a Public Goods Game (Mischkowski & Glöckner, 2016). Participants with a more

prosocial value orientation were faster when contributing more to the public good, whereas no response time relationships were found for participants in the pro-self range of the scale. As such, the time it takes to cooperate reflects the relative value of cooperation in the current environment or the social preferences of individuals.

Taken together, these findings are difficult to reconcile with the intuitive cooperation hypothesis (or, in fact, either of the two dual-process accounts). They reveal that cooperation itself is neither fast nor slow. Instead, the ease with which people cooperate is determined by the value attached to cooperation. The value of cooperation can vary across individuals or situations, but in each case it shapes how quickly people make cooperative decisions.

4.1. Self-control and the dlPFC in cooperation

Other evidence for the utility of value-based models of cooperation comes from the field of social neuroscience. A common view of the dorsolateral prefrontal cortex (dlPFC) during choice is that it implements deliberative attempts to override intuitive processes (Satpute & Lieberman, 2006). For example, the dlPFC has been implicated in overriding affective intuitions during utilitarian moral judgments (Greene, Sommerville, Nystrom, Darley, & Cohen, 2001). According to this view, the dlPFC should be associated with either suppressing or enhancing cooperation (depending on whether the dual-process theory of cooperation argues that cooperation requires intuition or deliberation, respectively).

In contrast to the intuitive cooperation hypothesis, we recently found evidence that the dlPFC is not necessary for inhibiting cooperation by examining behavior in the Public Goods Game in patients with brain lesions (Wills et al., 2018). Specifically, thirty-

seven patients with brain lesions due to a variety of medical conditions were recruited from the New York University Patient Registry for the Study of Perception, Emotion and Cognition. Patients were classified into different groups reflecting the primary location of their tissue damage. In total, 10 patients were classified as having frontal lobe damage extending into vmPFC or dlPFC, 16 patients were found to have anterior temporal lobe (ATL) resections (involving the amygdala and hippocampus), and 11 were classified as brain-damaged comparison (BDC) patients with mixed lobar (frontal, parietal or temporal) lesions. Among the 10 frontal patients, two had primarily dlPFC damage, 2 had primarily vmPFC damage, while 6 had damage to both regions. Degree of damage to each sub-region was treated as a continuous regressor in our analyses.

Participants played twenty one-shot public goods games with anonymous partners whose behavior had been pre-recorded. The task was, however, fully incentivized and without deception. The results revealed that the likelihood of cooperation fell with increased dlPFC damage (see **Figure 4)—**with zero cooperation among the patients with the largest damage to this region**.** In contrast, the likelihood of cooperation increased with increased vmPFC damage. Together, those findings are incompatible with a view of the dlPFC being a substrate for regulating or overriding intuitive cooperation. One interpretation of the data could be to support a view of humans being intuitively selfish with controlled processes intervening to promote prosocial decisions. However, the data are also compatible with value-based decision-making.

Recent research suggests that the dlPFC likely performs multiple functions in human decision-making. Evidence from studies of value-based decision-making using

primary reinforcers (e.g., juice and erotic images) has implicated the dlPFC in integrating value signals computed in the vmPFC (Hare, Schultz, Camerer, O'Doherty & Rangel, 2011; Domenech, Redouté, Koechlin, & Dreher, 2018). For instance, the source of value signals has been localized as inputs from the vmPFC to the dlPFC (Domenech et al., 2018). Another important function of the dlPFC is modulating subjective value representation through executive control–which allows goals to shape decisions. For instance, the dlPFC shifts the weights from attributes which are most salient to the decision-maker (e.g., money for self among selfish people) to other attributes (e.g., money for others; Hutcherson & Tusche, 2020). In the domain of cooperation, the dlPFC appears to shift the valuation of cooperation depending on context to promote goal-specific behavior (Buckholtz & Marois, 2012; Carlson & Crockett, 2018; Tusche & Hutcherson, 2018; Hutcherson & Tusche, 2020). This process, in turn, alters the value placed on selfish or cooperative decisions.

A recent neuroimaging study is consistent with this interpretation of the role of the dlPFC in cooperation (Hackel et al., 2020). In this experiment, students played a series of one-shot public goods games ostensibly with other university students while undergoing fMRI. Participants were told that the other university students they would be playing with were either entirely from a prosocial school (where 70% of students cooperated) or an antisocial school (where 30% of students cooperated). Participants then alternated every 25 trials between playing with students from each school.[2] Earnings across all trials were averaged and paid to students after the study. Prosocial tendencies were computed for each participant by computing their average level of

[2] At the outset, participants were told that one school was more likely to cooperate than the other, but were not told which school was which.

cooperation (i.e. the proportion of trials they decided to give). Those who were more cooperative in their choices showed greater vmPFC activity when cooperating compared to more selfish individuals (see **Figure 5**). In the more prosocial group, dlPFC activity was associated with making selfish choices and with increased dlPFC-vmPFC connectivity. In other words, the vmPFC appeared to reflect the value individuals place on cooperation and the dlPFC appeared to be involved in modulating the impact of these value computations on decisions to cooperate.

5. Cognitive Processes

In this section, we discuss several additional cognitive processes, apart from self-control discussed above, that contribute to determining the value of cooperation (see Figure 1). Of these, representing other minds is highly recognized in the literature while attention and memory are processes which are less commonly studied in cooperation research. The focus on attention and memory comes as a natural extension of adopting a value-based framework. Should these processes, or the social inputs discussed in the next section, not meaningfully contribute to cooperative decision-making this would constitute reasons to reformulate how we expect domain-general processes to contribute to cooperative decision-making within our framework. Finally, the processes discussed in this and the previous section are not meant to be an exhaustive enumeration of inputs but rather serve as a starting point for understanding current and future research.

5.1. Representing other minds

Cooperation crucially depends on the actions of other people, hence expectations about others' behavior and what social norms are in place plays a large

part in determining the value of cooperation. Most people are conditional cooperators – meaning they are more likely to cooperate if others cooperate and vice versa (Fehr & Schurtenberger, 2018; Gill, Packer, & Van Bavel, 2013). Because of this, manipulating people's expectations about what other people consider to be fair (i.e., a social norm of fairness) affects the likelihood of accepting unfair offers (Chang & Sanfey, 2013; Xiang, Lohrenz, & Montague, 2013). Activity in the anterior cingulate cortex and anterior insula has been associated with updating and learning expectations of social behavior.

The insula is also associated with representing social norms. For example, patients with insula lesions update their internal norms slower in response to feedback (Gu et al., 2015). Consistent with its role in representing norms, insula activation has been found to be correlated with receiving unfair offers and with consequent rejection of such offers (Sanfey, Rilling, Aronson, Nystrom, & Cohen, 2003). Insula activity also tracks individual differences in aversion to inequity–a norm of fairness (Hsu, Anen, & Quartz, 2008). In sum, cooperation decisions are dynamic and involve representing and updating norms, and this expectation arises from activity in multiple brain regions.

The value of cooperation depends on variation in individuals' social preferences. The social cognition network, which includes the temporoparietal junction (TPJ) and the adjacent posterior temporal sulcus (pSTS), aids representations of others' minds, goals and intentions (Parkinson, Kleinbaum, & Wheatley, 2017). Inferring trustworthy intentions and goals of fellow group members is fundamental to cooperation decisions (Yamagishi & Kiyonari, 2000). Functional and anatomical variation in the social cognition network has been associated with individual differences in social preferences (Morishima, Schunk, Bruhin, Ruff, & Fehr, 2012) and activity in the pSTS correlates with

willingness-to-give when choosing charitable donations (Hare, Camerer, Knoepfle, O'Doherty, & Rangel, 2010). Functional connectivity between TPJ and vmPFC may indicate modulation of cooperative values through social preferences (Strombach et al., 2015). Hence, cooperation decisions appear to be modulated by social information processed through these brain regions.

5.2 Attention

The value of options during choice are modulated by attention. Attention to options influences the likelihood of choosing them, such that more attended options are more likely to be chosen (Krajbich et al., 2010; Shimojo, Simion, Shimojo, & Scheier, 2003; Smith & Krajbich, 2018, 2019). The rate of evidence accumulation for decision options depends on what option a decision-maker is attending to, which can account for the patterns of visual fixations, response time, and choices made in value-based decision tasks (Krajbich et al., 2010; Thomas, Molter, Krajbich, Heekeren, & Mohr, 2019). Value signals in the striatum and vmPFC have been found to be modulated by relative fixation time to different options (Lim, O'Doherty, & Rangel, 2011). This relationship between attention, choices, and response times also extends to choices based on social and moral preferences (Pärnamets et al., 2014; Smith & Krajbich, 2018; Teoh et al., 2020).

Attention causally influences the outcome of decisions by modifying value. Forcing attention to options by manipulating presentation times can bias decisions between faces (Shimojo et al., 2003) and snack foods (Armel, Beaumel, & Rangel, 2008). Causal effects of attention on choices have been extended into the moral domain, revealing that manipulating attention has a clear influence on a wide range of

value judgments (Pärnamets et al., 2015; Ghaffari & Fiedler, 2018; Falandays & Spivey, 2020; Fosgaard, Jacobsen, & Street, 2021; but see Newell & Le Pelley, 2018). In one study, for instance, people were asked to respond to morally charged items such as “Is murder justifiable?”, and were given two response options, “sometimes justifiable” and “never justifiable” (Pärnamets et al., 2015). Based on visual attention, determined by their looking times towards the options, participants’ were prompted to make a decision. Triggering decisions based on visual attention subtly biased what moral judgment they made. Thus, attention appears to directly influence moral decision-making.

Little research has examined how attentional dynamics influence decisions to cooperate in larger, multiplayer cooperative dilemmas. According to our framework, it is likely that the value of cooperation is similarly affected by decision makers' fluctuations in attention. For instance, several studies have found that people’s social preferences influence what information they attend to when viewing social dilemmas (Fiedler, Glöckner, Nicklisch, & Dickert, 2013; Jiang, Potters, & Funaki, 2016). In addition, instructing people to attend to different attributes appears to alter the weights given to self and other related outcomes in modified dictator games (Tusche & Hutcherson, 2018). Recent computational modeling suggests that social preferences influence early allocation of attention, and consequent changes in dwell times to self or other related information alter what choices people ultimately make (Teoh et al., 2020). Thus, attention may play a key role in group based cooperation.

Past research has found that the presence of consistent contributors–people who contribute to the public good repeatedly–can increase cooperation rates among fellow group members (Fowler & Christakis, 2010; Weber & Murnighan, 2008), and it is likely

that this effect is moderated by attention to those consistent contributors. For instance, we have found that attention to social partners who cooperated on previous rounds of a public goods game was associated with cooperation on the next round (Pärnamets, Gill, Packer, & Van Bavel, 2021). In this study, 96 participants were recruited from the local participant pool and played a repeated, 28 trial public-goods game against three pre-programmed avatars who they were led to believe were actual participants. One of the avatars was a consistent contributor (Weber & Murnighan, 2008) and always contributed, while the other two contributed variably. While the participants played the game their eye-movements were recorded using a remote eye-tracker. We investigated how patterns of attention during the outcome phase of the game - when the choices of the avatars during each round were revealed to the participant - affected their likelihood to cooperate on the subsequent trial. Our analyses revealed that, contrary to our initial expectations, increased attention to the consistent contributor did not impact cooperation. However, variation in trial-by-trial attention to the other two avatars did impact cooperation, conditional on those avatars having cooperated in the previous round of play. This suggests that the value of cooperation is modulated by attentional shifts in the immediate social environment. More generally, if attention to partners in social dilemmas systematically correlates with cooperation, then investigating the causal structure of this relationship may be useful for creating interventions aimed at boosting cooperation. We believe this is an important, but understudied topic in the domain of social cooperation.

5.3. Memory

In a value-based framework, the values of actions when facing cooperative dilemmas are assumed to be constructed based on multiple inputs to the core valuation systems. Value is assumed to be acquired by learning, whether direct or social (FeldmanHall & Dunsmoor, 2019; Olsson, Knapska, & Lindström, 2020; Staats & Staats, 1958). Recently, the memory system–and specifically the hippocampus–has increasingly been implicated as playing an important role in the construction of values during decision making by sampling episodic memories related to the current decision context (Shadlen & Shohamy, 2016). For example, changes in connectivity between hippocampus and the core valuation areas of the striatum and vmPFC affects value-based decisions for monetary rewards (Shadlen & Shohamy, 2016; Wimmer & Shohamy, 2012), and patients with hippocampal lesions make more random decisions (Bakkour et al., 2019), and are worse acting according to internal models of the decision space (Vikbladh et al., 2019). This suggests that understanding the role of memory will be critical for understanding group cooperation.

In our view, it is highly likely that the hippocampus is key for fostering decisions to cooperate. For instance, hippocampal activity has been found to support adaptive social decision-making based on past encounters with social partners in the context of a dictator game (FeldmanHall, Montez, Phelps, Davachi, & Murty, 2021). Deciding to cooperate may involve recalling both a generalized value of being prosocial or not, but also, past episodes involving the people in the current cooperative context. If these memories are salient, they can influence decisions. Indeed, studies have found that imagining prosocial acts through episodic simulation increases prosociality (Gaesser & Schachter, 2014). This is still an emerging area of research in social decision-making

and more work should examine the role of cognitive processes–like attention and memory–on group cooperation.

6. Psychological and contextual influences

6.1. Social Norms

Decisions to cooperate also hinge on the social context of the decision-maker. Specifically, cooperation hinges on *who* people are cooperating with and on their beliefs about whether these other people will be cooperative in return (see Yamagishi & Kiyonari, 2000). This is one reason why groups and social identity may play such a central role in cooperative decisions–they provide a clear signal that others will reciprocate cooperation in future interactions. In addition, through repeated interactions and experience over time in a given group or culture, people may form expectations that guide their cooperative behavior (Hackel et al., 2020). As such, social norms and group identities play a role in cooperation (Kramer & Brewer, 1984; Van Bavel & Packer, 2021) and these influences can cascade as people join new groups and shape norms of cooperation (Fowler & Christakis, 2010). Although these differences emerge in small groups and communities, it is unclear whether these scale to national level differences in cooperation (Sparado et al., 2022).

Within groups, social norms provide shared standards of behavior based on common beliefs about how in-group members should act in a given situation (Cialdini, 1991; Fehr & Fischbacher, 2004; Fehr & Schurtenburger, 2018). More specifically norms have been defined as informal behavioral rules that people follow conditional on two criteria: (i) if they think a majority of others in their community follow that rule and (ii) if they think a majority of others in their community also think the rule should be followed

and are willing to impose sanctions for those who do not follow the rule (Bicchieri, 2005a; Cialdini, 1991). Thus, people have lay beliefs about what is typically done – a descriptive norm – and what ought to be done – an injunctive norm. Together, these social perceptions shape decisions to cooperate with others.

At times, people appear to be more influenced by descriptive norms. For example, in a field study conducted on energy use in 287 households in the USA, participants were provided with descriptive norm information about their energy use relative to that of neighbors. The descriptive norm information influenced energy consumption for all households: households who were consuming more energy than average began consuming less, and households who were consuming less energy than average began consuming more[3] (Schultz et al., 2007). Thus people's real-world cooperative behavior was influenced by descriptive normative information.

However at other times, people appear to be more influenced by injunctive norms. For example, in a study that directly compared the effects of providing descriptive (e.g., *Most Player 1s give $0.20 or more to Player 2*) versus injunctive norm information (e.g., *It is suggested that Player 1 give $0.20 or more to Player 2*) in the context of a Dictator Game, found that injunctive norms increased compliance relative to the control condition. In other words, telling participants that they ought to be somewhat cooperative in an economic game led to participants actually being more cooperative (Raihani & McAuliffe, 2014).

A norm of conditional cooperation is commonly identified as the most important social norm explaining behavior in social dilemmas (Fischbacher, Gächter, & Fehr,

[3] This latter unintended effect could be removed by adding injunctive norm information in the form of a smiling or sad emoticon next to the information about the householder's energy use.

2001; Fehr & Schurtenberger, 2018). Under conditional cooperation people cooperate depending on their expectation of others' level of cooperation. Some work examined cooperation on a Prisoner's Dilemma task with actual female prisoners and compared rates of cooperation to female undergraduate students (Khadjavi & Lange, 2013). In a simultaneous Prisoner's Dilemma (where both participants make a choice about whether to cooperate or defect at the same time), 56% of inmates chose to cooperate whereas only 37% of students chose to cooperate. This suggests that prisoners may be more cooperative with one another, potentially due to reciprocity and punishment norms that are more salient among prisoners. However, the rates of conditional cooperation in a sequential Prisoner's Dilemma between prisoners and students were similar, around 60%. Thus, once a partner has already signaled that they are cooperative, it may set the stage for future cooperation, altering any pre-existing social norms.

When conditional cooperators are able to dynamically choose who they cooperate with, it can reinforce certain norms of cooperation while excluding untrustworthy defectors (DeSteno, 2014; Wang et al., 2012). For example, in one online network study, groups of 24 participants played an iterated prisoner's dilemma game and were allowed to propose and remove links to players of their choosing at different times. Participants were able to update their cooperation strategy (cooperate vs. defect). Some participants were able to update their strategy after each round (most dynamic), whereas others were only able to update their strategy after every four rounds, and still others were only able to update once across the 12 rounds (least dynamic). Moreover, participants were given the opportunity to update who they played with. Some participants were allowed to update five social links each time (most

dynamic), others were allowed to update three social links each time, and others were only allowed to update one social link each time (least dynamic). Across a variety of different conditions, dynamic partner updating increased the level of cooperation and the average payoffs to players (Wang et al., 2012). This suggests that people are sensitive to defectors and will punish defectors by severing ties with them to selectively cooperate with more cooperative partners. This explicit conditional cooperation can build a stronger social norm for cooperation.

Among conditional cooperators, some are more responsive to certain aspects of social norms than others. For instance, one study found that nearly 12% of participants cooperated primarily because they thought others would cooperate (first-order empirical expectations); 14% of participants cooperated primarily because they thought others would think everyone ought to cooperate (second-order normative expectations), and 11% cooperated for both reasons (Szekely et al., 2021). In addition, nearly 27% of participants were strategic cooperators, decreasing their contribution when others increased, whereas 37% of participants were unconditional cooperators or consistent contributors who did not change their behavior in response to expectations (see Gill, Packer, & Van Bavel, 2013). Thus, roughly two-thirds of participants in the sample were influenced by social norms.

Experiments have also found that social norms causally affect cooperation. For instance, researchers who manipulated sign-postings in public spaces highlighting norms against littering found that these norms reduced littering (Kallgren, Reno, & Cialdini, 2000). In a recent 30-day social dilemma experiment (Szekely et al., 2021), researchers also observed causal effects of social norms on contributions to a common

goal. Specifically, people contributed more to a public pot both when they thought that their group would contribute more as well as when they thought their group would think everyone should contribute more. Moreover, at the end of the 30 days, people punished people who contributed less and expected others in their group to punish individuals who contributed less. Thus, perceived group norms play a powerful role in determining both cooperation and responses to non-cooperators.

A value-based approach can explain how social norms influence cooperation. This approach takes into account variability in people's prosocial tendencies and sensitivity to group norms. As a result, individual differences in prosociality and situational features shape the value-computation of cooperation (Van Lange et al., 1992, 2013). There is a growing body of evidence that prosocial tendencies and social norms interact to guide cooperation. As we noted above, descriptive norm manipulation can alter the social norms of cooperation within a single experiment (Hackel et al., 2020). Specifically, people who were highly sensitive to this shifting context cooperated nearly twice as often with members of the more cooperative group. People were also able to shift back and forth between cooperative and selfish strategies to reflect the norms of the group they were playing with at any given moment. This underscores the dynamic nature of social attunement to social norms.

In Hackel et al.'s (2020) study, the impact of social norms suggests that people successfully detected the social norm differences (i.e., noticing that students from one group cooperated more often than students from another group). Norm detection was not evenly distributed, but it was associated with cooperative behavior. People who were better at encoding greater cooperation from the prosocial school were also more

cooperative toward members of the prosocial school. The effect of the social norms was also seen in the participants' neural activity. Specifically, connectivity between the dlPFC and the vmPFC increased when participants made decisions that deviated from the norm of the school they currently were playing with (i.e., when cooperating with the antisocial school or defecting with the prosocial school). These results suggest that the brain is tracking social norms and computing decision conflict (e.g., whether specific cooperative decisions are aligned with one's prosocial tendencies), consistent with a value-based approach.

Social norms of cooperation also vary cross-culturally. For example, there was considerable variability in the average level of cooperation (measured via contributions to a public goods game) across the 21 worldwide research sites involved in the registered replication of the intuitive cooperation hypothesis we mentioned above (see **Figure 6**). For instance, the cooperation rate in the same task was nearly twice as common in Budapest, Hungary, than at our lab in New York City, USA. These data suggest that normative differences in group-based cooperation vary a great deal across varying geographies and group identities.

The local base rates of cooperation in a community can shape the value people place on the decision to cooperate or defect by affecting expectations. If most other players are typically cooperative (e.g., Budapest, Hungary), people may be aware of that social norm and default to cooperating as a maximizing decision strategy. Conversely, if most other players are typically less cooperative (e.g., New York, USA), people may default to selfishness as a maximizing strategy. As such, a value-based model would predict an association between the effectiveness of a time-pressure

manipulation–which elicits more automatic decision making–and local norms. Specifically, forcing rushed decisions should increase cooperation in cooperative environments but reduce cooperation in selfish environments. Indeed, a re-analysis of the global registered replication data revealed that time pressure had a greater boost on cooperation at locations with more cooperative participants but reduced cooperation at sites with less cooperative participants (Pärnamets, Wills, & Van Bavel, 2021). These results are consistent with a value-based account which considers individual differences in norm sensitivity, prosocial tendencies, and existing social norms.

6.2. Social identities

While prosocial tendencies and social norms shape the valuation of cooperation, group identities can also impact cooperation. When we trust others in our group we are more likely to cooperate with them, which is a form of in-group favoritism (Robinson & Barker, 2017). People cooperate more with in-group than out-group targets (Brewer & Kramer, 1986; Yamagishi & Kiyonari, 2000; Chen & Li, 2009; Charness, Cobo-Reyes & Jiménez, 2014) and trust plays a central role in cooperation within groups (e.g., Dawes, 1980; Edney, 1980; Foddy, Platow, & Yamagishi, 2009; Kramer & Brewer, 1984). Research has shown that ingroups are favored in cooperation both for naturally occurring groups (Bernhard, Fehr, & Fischbacher, 2006; Goette, Huffman, & Meier, 2006) and when artificial groups are created in the laboratory using variations of the minimal group paradigm (Yamagishi & Kiyonari, 2000; Chen & Li, 2009; Charness et al., 2014).

One study examined several methods of inducing group membership in a large number of economic games (Chen & Li, 2009). They found that participants were more

generous towards ingroup members when making monetary allocations between themselves and another person, but also that when participants were matched with ingroup partners the overall rate of socially welfare maximizing choices—cooperative choices—increased. Generosity was higher and envy was lower when participants made decisions between themselves and an ingroup partner compared to when deciding about an outgroup partner. Finally, ingroup bias was unaffected by whether minimal groups were constructed based on participants' preferences or randomly assigned, but they were strengthened if group members were given additional opportunities to communicate in unrelated tasks prior to the economic games. Moreover, when members within a group unanimously make promises to cooperate it signals a group identity (Dawes, Van De Kragt, & Orbell, 1988). Allowing group members to communicate during social dilemmas drastically increases rates of cooperation in part due to promise-making and the fact that group norms can be made salient (Bicchieri, 2002; 2005b).[4] As a result, norms feed group identities, which can impact cooperative behavior.

Making shared group identities salient can play a powerful role inducing cooperation (De Cremer & van Vugt, 1999). For example, Reinero et al. (2021) recruited groups of four to the lab and randomly assigned some groups to work together on problem-solving tasks as a collaborative team and other groups to work on the same set of problem-solving tasks but as competitive individuals. Teams formed a team name and built rapport through a behavioral synchrony manipulation, whereas the group of individuals created their own individual code names. Thus, the key difference between

[4] Bichhieri argues that social norms explain cooperative behavior more so than group identity.

the team and individual condition was a shift in the psychological experience from one of a collective team (social identity) to one of a competitive individual (personal identity).

People in the team condition were more highly identified than people in the individual condition. Importantly, this shared identity manipulation increased cooperation: people in the team condition contributed more money to their group's public pot than those in the individual condition. While many participants were highly cooperative, contributing the full $10 to their group's public pot, 74% of people in the team condition contributed the full amount compared to only 51% in the individual condition. Moreover, group identification was associated with cooperation, such that people who were more highly identified with their group were also more cooperative regardless of condition. Indeed, other work has found that people's identification with the world as a whole (a global social identity) predicts behavioral contributions to a global public good beyond what is predicted from expectations about what other people are likely to contribute (Buchan et al., 2011). These results point to the fact that group identity can be an important determinant of cooperation.

While group identities clearly impact cooperation, the effect of groups is not always unambiguously positive. For example, several studies have found that ingroup norm violators are treated more leniently compared to outgroup norm violators (Chen & Li, 2009; Bernhard et al., 2006; but see McLeish & Oxoboy, 2007). Such behavior is usually explained by appealing to reciprocity expectations being stronger in ingroup situations, but nevertheless highlight the complex motives governing both cooperation specifically and group interaction generally. Ingroup preferences in cooperation can, in

the aggregate, also produce significant disparities between groups—especially if there are inequitable resource distributions to start with.

6.3. Social Networks

Cooperation becomes more complex as the group size grows. Understanding how psychological phenomena scales from individuals to dyads to social networks is central to understanding all sorts of collective-level behaviors (Vlasceanu, Enz, & Coman, 2018). For example, tracking reputational information is crucial in social network based cooperation and relies heavily on the ability to remember who did what in the past (or what you've heard about people's behavior through gossip). One of the hypothesized limits on network size is the ability to track exponentially increasing interrelationships between people as networks grow (Dunbar, 1998). As such, attentional and memory processes (tracking who did what) can be altered in larger social networks, which can influence the value-computation of cooperation. For instance, compared to dyads, groups exhibit different patterns of non-verbal behaviors (e.g., eye gaze) and communication (Solano & Dunnam, 1985; Herrera, Novick, Jan, & Traum, 2011), as well as different motivations to trust and cooperate with others (Zhou & Zhang, 2006; Wildschut & Insko, 2007; Pereda, Capraro, & Sánchez, 2019).

While larger social networks may make it more difficult to track reputational information of everyone, there is also experimental evidence that increasing group size (e.g., 5 vs. 40) boosts cooperation in a public goods game (Pereda et al., 2019). As we noted above, cooperation can cascade through social networks as people are influenced by others' cooperative behavior and "pay it forward" to new groups of people (Fowler & Christakis, 2010). In this research, participants played six repeated one-shot

public goods games in groups of 4, each game with a different set of group members and all members were anonymous. The design of the study allowed the researchers to examine the degree to which an individual's cooperative behavior was influenced by their fellow group members' behaviors in preceding rounds (influence of directly connected individuals). In addition, the researchers were able to analyze the association between indirectly connected participants to identify whether such effects spread from person to person to person. For example, Annie may influence Jack, who in turn influences Matty, even though Matty did not interact with Annie or observe her behavior. The researchers found that not only was influence present among directly connected participants, but also that this influence persisted for up to three degrees of separation (from person to person to person to person). Thus, cooperation spread through a network, tripling in impact via other subjects who were directly or indirectly influenced to contribute more as a consequence.

Considering how the value-computation process for cooperation is altered when moving beyond dyads to larger groups and social networks is crucial. Not only do individual-level cognitive processes, such as attention and memory, become impacted, but so do broader collective-level aspects such as the degree of social influence within a network, perceived and dynamic norms, and collective identities.

6.4. Social Institutions

Trust production is crucial for fostering cooperation (Zucker, 1986). We have already discussed two forms of trust production above: the trust and resulting cooperation that develops from experience with and knowledge about individuals, and trust based on social identities. The third form of trust production is institution-based, in

which formal mechanisms or processes are used to foster trust (and that do not rely on personal characteristics, a history of exchange, or identity characteristics). At the societal level, trust-supporting institutions include governments, corporate structures, criminal and civil legal systems, contract law and property rights, insurance, and stock markets. When they function effectively, institutions allow for broader cooperation, helping people extend trust beyond other people they know or know of and, crucially, also beyond the boundaries of their in-groups (Fabbri, 2022; Hruschka & Henrich, 2013; Rothstein & Stolle, 2008; Zucker, 1986). Conversely, when these sorts of structures do not function well, “institutional distrust strips away a basic sense that one is protected from exploitation, thus reducing trust between strangers, which is at the core of functioning societies” (van Prooijen, Spadaro, & Wang, 2022).

When strangers with different cultural backgrounds have to interact, it often lacks the interpersonal or group-level trust necessary for cooperation. For instance, reliance on tightly-knit social networks, where everyone knows everyone, is often impossible in larger, more diverse environments. Communities can compensate by relying more on group-based trust. For example, banks may loan money primarily within separate kin or ethnic groups (Zucker, 1986). However, the disruption of homogeneous social networks, combined with the increasing need to cooperate across group boundaries creates incentives to develop and participate in broader sets of institutions. Institutions can facilitate cooperation and individuals prefer institutions that help regulate interactions and foster trust.

People often may seek to build institutions embodying principles, norms, rules, or procedures that foster group-based cooperation. In turn, these institutions shape

decisions by altering the value people place on cooperative decisions. One study, for instance, examined these institutional and psychological dynamics over 30 rounds of a public goods game (Gürerk, Irlenbusch & Rockenbach, 2006). Every round had three stages. First, participants chose whether they wanted to play that round with or without a “sanctioning institution” that would provide a means of rewarding or punishing other players based on their behavior in the game. Second, they played the public goods game with (and *only* with) other participants who had selected the same institutional structure for that round. After making their decisions (to contribute to the common pool), they then saw how much everyone else in their institutional context had contributed. Third, participants who had opted to play the round with a sanctioning institution could choose, for a price, to punish or reward other players.

Initially, only about a third of participants chose to play with a sanctioning institution. Cooperation was, however, significantly higher among those players. Forty-eight percent of participants who played with an institution on the first round gave at least 75% of their endowment to the collective, and a mere 16% were “free-riders”, giving only a quarter or less. In contrast, among participants who opted to play without a sanctioning institution, only about 11% gave 75% or more of their endowment and 43% were free-riders. Thus, having a sanctioning institution significantly boosted cooperation.

Over the 30 rounds, participants' preferences for institutional structures increased dramatically. By the end, the overwhelming majority of players (93%) opted to play the public goods game in the presence of a sanctioning institution and this had profound benefits for cooperation. On average throughout the entire procedure, participants

playing with a sanctioning institution contributed 91% of their endowment, while participants playing without an institution contributed a measly 14%. In this experiment, the trust-producing institution is one that allows individuals to reward and punish each other directly, thus altering incentives for cooperation. In many contexts, of course, institutions are much more impersonal than this, operating external to the interacting parties themselves.

Recognizing the wide variety of procedures, rules, and structures that can serve organizing functions (see for example Ostrom, 1990), institutions can be defined as "systems of established and prevalent social rules that structure social interactions" (Hodgson, 2006). Mechanisms for facilitating cooperation, including monitoring and enforcement processes, are more institutional in nature to the degree that they are *formalized*, such that rules are codified and explicit, and *standardized*, such that they are applied impartially across people and situations (Zucker, 1986). From the perspective of our model, these features ensure fair *outcomes* and increase the *probability* of reciprocity. To the extent that people perceive the existence of effective trust-facilitating institutions, they should be more willing to trust others, including strangers. There should be an association between positive perceptions of societal institutions and cooperation within those societies.

Indeed, trust in institutions like the police, legal system, and government is positively associated with interpersonal trust across many nations. For example, institutional trust correlates with generalized trust in nationally representative data from 16 European countries (Spadaro, Gangl, Van Prooijen, Van Lange, & Mosso, 2020). Similarly, trust in societal institutions predicts the endorsement of cooperation (over

cheating) in real-world social dilemmas, including not avoiding fares on public transportation, not claiming benefits to which you are not entitled, and not cheating on taxes (Irwin, 2009). This relationship held for both individualist societies (including the United States, United Kingdom, and Australia) and collectivist societies (including Venezuela, Indonesia, Taiwan, and China, among others).

These sorts of findings are, of course, correlational and a plausible argument can be made for reverse causal relationships. For example, while effective institutions may facilitate trust and cooperation within societies, it may also be the case that societies with higher levels of trust among the population are able to build and maintain more effective institutions. At the heart of the debate over how social trust is created and maintained lie competing views framing trust as culturally or institutionally determined (Nannestad, Svendsen, Dinesen, & Sønderskov, 2014). On the former view, trust is socialized in cultural groups (e.g., through social norms of reciprocity) and can therefore be resistant to quick changes. Support for this comes from studies on second and third-generation immigrants primarily in the United States, revealing that trust levels in immigrants' country of origin correlated with their social trust even after several decades in the US (Dinesen, 2013; Dinesen & Sønderskov, 2018). However, studies from other countries, primarily in Europe, have found greater adaptation of immigrants' trust to the levels in their new home societies and that these effects are larger than the cultural residual (Dinesen & Sønderskov, 2018). Data from Denmark, in which citizens were surveyed multiple times over 18 years, revealed that institutional trust predicted changes in interpersonal, generalized trust over time (Sønderskov & Dinesen, 2016).

An ambitious field study, which took advantage of land rights reforms in Benin, found a clear causal impact of institutions on national cooperation (Fabbri, 2022). Between 2009 and 2011, the Benin government enacted land reforms that replaced informal social norms regarding individuals' rights to land with "formalized land rights enforceable by state courts". Crucially, these reforms were implemented in a randomized control trial in local villages. People from villages that had undergone institutional land rights reform contributed significantly more to their country account—a broad public good–than people from villages where land rights had not been institutionalized. They concluded that, "awarding formalized land rights enforceable in the state's courts…relaxes individuals' dependency on the protection offered by the village network and increases cooperation with strangers from other villages". Thus, introducing an institution allowed for greater cooperation with out-group members.

6.5. Substitutability of trust production

The three modes of trust production we discussed above are "to some extent substitutable" (Zucker, 1986), which suggests that the presence of effective institutions may reduce people's reliance on social network connections and shared group identities. As a result, in-group favoring biases may be lower in locations or contexts where people trust institutions to help regulate interactions. Consistent with this, a study of in-group favoritism across 122 societies found that this pattern bias was negatively associated with government effectiveness (indexed by the World Bank and controlling for a host of variables, including pathogen stress, religion, and inequality; Hruschka & Henrich, 2013). Similarly, people in societies with less effective institutions exhibited a greater preference for their in-group than people in societies with more effective

institutions (Hruschka et al., 2014). These cross-cultural studies underscore the value of building trustworthy institutions to erode the impact of social identities in guiding cooperation.

Trust-facilitating institutions might also reduce implicit intergroup bias (Lin & Packer, 2017). In this research, White participants learned that they would play a series of trust games with White and Black interaction partners, and half were informed that there would be a monitoring institution present to sanction non-cooperation. Prior to playing the trust games, participants completed an evaluative priming task with White and Black faces. Participants who expected to play trust games with same and other-race partners without a sanctioning institution exhibited a standard pattern of racial bias with a preference for White over Black faces (e.g., Fazio et al., 1995). However, racial bias was reduced among people who anticipated playing trust games with a sanctioning institution. Effective trust-producing institutions may, then, be an important guard against intergroup biases, as well as nepotism that favors friends and family over strangers.

It is concerning that trust in important societal institutions has been declining in many nations. It also raises the question of whether institutional decay is a causal contributor to rising levels of intergroup conflict, hate, and nationalism as people turn inward to secure trust and cooperation (van Prooijen et al., 2022). Several psychological mechanisms might contribute to the positive influence of institutions on cooperation with strangers beyond known network contacts and in-group boundaries. For example, institutions may increase cooperation in a strategic fashion by providing “assurance”.

People interacting in the presence of a sanctioning institution can assume that others have an incentive to be cooperative–because if they are not, they will be punished.

While assurance mechanisms increase cooperation, they do not actually produce trust. However, this conclusion depends on how one defines the terms. One approach to trust focuses on people's beliefs that other people are interested in their outcomes, or at least in pursuing mutual benefit, and are not entirely self-interested. But another approach suggests that trust occurs when people can assume that interactions will transpire as they expect them to, following predictable scripts for behavior. While behaving altruistically at church should increase trust, being altruistic while conducting banking transactions would be unusual, raising eyebrows and possibly disrupting trust. When operating as assurance mechanisms and thereby providing stability and predictability to interactions, institutions may be increasing the latter more than the former form of trust.

However, institutions may also increase trust and cooperation through less strategic and more generalized mechanisms. For instance, "institutions initially influence trustworthy interactions through external rules, but after repeated successful interactions, individual's would not rely on mere assurance anymore, generalizing their beliefs about others' benevolence to other settings" (Spadaro et al., 2020). As such, trust in institutions has an indirect effect on interpersonal trust via greater feelings of security, such as feeling protected by public institutions.

Similarly, institutions that increase accountability and enforcement create top-down incentives to cooperate, which can lead people to adopt heuristics prescribing cooperation in a wide variety of situations (Stagnaro, Arechar, & Rand, 2017). For

instance, institutions are associated with more *prosocial* behavior, not just greater cooperation when it is strategically advantageous. Specifically, people who reported greater confidence in the police and the courts gave more money to a partner in a one-shot dictator game. Effective institutions can also produce trust and cooperation by "providing models for group norms and values" (van Prooijen et al., 2022). People may, for example, draw conclusions about normative or appropriate behaviors by observing the actions of actors who represent or embody important institutions–such as police officers, judges, or political leaders. In corrupt systems, people may observe others behaving corruptly and conclude that "in order to get what one needs in life, one must be engaged in various forms of corruption".

We have discussed institutions as a broad class of relatively formalized and standardized rules, procedures, etc. But institutions come in different forms with varying goals and functions, and different types may have distinct effects on trust and cooperation. At the societal level, it is important to distinguish between political institutions involved in *representation*, including political parties, parliaments, houses of congress, etc., and institutions involved in *implementation,* including career civil services, police, and judicial systems (Rothstein & Stolle, 2008). Trust in representational institutions is generally partisan–determined largely by group identity–and that perceptions of their effectiveness will ebb and flow in different segments of the population depending on who currently holds power. People will generally experience more trust in these institutions when the political agents they support have control.

In contrast, implementational institutions are expected to be impartial and non-partisan. Trust in them should thus be more stable, grounded in their actual institutional

efficacy rather than *who* currently holds political office. As such, generalized trust is often more strongly associated with trust in implementational than representative institutions. For instance, in a sample of 57 countries, national levels of trust in institutions including the police, legal system, and army were positively associated with generalized trust. In contrast, trust in representational or partisan institutions was *not* associated with generalized trust (Rothstein & Stolle, 2008).

Further differentiation of institutions is important for understanding specific types of cooperative behavior. During the COVID-19 pandemic, for example, confidence in science as a type of institution has proven to be an important predictor of engaging in health-protective behaviors. Longitudinal and representative data from 12 countries revealed that trust in scientists was consistently associated with support for societal measures like closing schools or imposing quarantines, compliance with social distancing recommendations and rules, and endorsement of vaccination (Algan, Cohen, Davoine, Foucault, & Stantcheva, 2021). Similarly, trust in science was related to vaccine confidence within and across 126 nations (Sturgis, Brunton-Smith, & Jackson, 2021). In contrast, these populations' trust in their governments was inconsistently and more weakly related to health-protective behaviors. While governmental trust was linked to greater compliance in some nations, it was actually associated with less compliance in countries like the United States and Brazil, where government and scientific authorities were frequently at odds with each other in terms of how to respond to the crisis. Likewise, science skepticism in the US (as indexed by county-level disbelief in human-caused global warming) was associated with less physical distancing early

during the pandemic, over and above effects of political partisanship, as well as local rates of infection and mortality (Brzezinski, Kecht, Van Dijcke, & Wright, 2021).

7. The future of cooperation

In the late 1960's, the American Institute of Planners invited a list of the brightest scholars and policy makers to predict the biggest issues facing humanity in the future. Although some of the responses now look absurd, such as worries about million-ton planes, many of the concerns were prescient. Over a half century earlier, they foresaw one of the major societal and global problems facing humanity today: catastrophic climate change. Tackling this problem will require group-based cooperation at a massive scale—far beyond what we have discussed thus far in the chapter.

By now, most scientists are aware that the earth is warming at an unsustainable rate due to human economic activity and it could have catastrophic consequences for humanity. Our only solution is to cooperate with people around the world on a set of policies and actions that will, collectively, stem the rising temperature and preserve the earth for future generations. Unfortunately, this will require a form of cooperation that is unprecedented in human history and beyond the scope of our normal institutions which are organized and funded by national governments.

As such, the cooperation necessary for addressing climate change, and other global problems, will require bringing different collectives together to cooperate on behalf of *future* generations. Unfortunately, when the rewards of self-interest are immediate and the benefits of cooperation are accrued by future generations or delayed by several decades it can lead to a marked decrease in cooperation (Jacquet et al., 2013). A recent review of the literature also suggests that more parochial identities may

provide a barrier to this form of cooperation—as attitudes towards preventing climate change are polarized in some of the world's biggest polluting nations (Doell, Parnamets, Harris, Hackel & Van Bavel, 2021; Hornsey, Harris, & Fielding, 2018). As such, the first barrier to overcome will require navigating partisanship within these target countries, including the USA, UK, Australia, etc. where right-wing parties are far less likely to accept the science on climate change. More work will need to tackle this particular challenge in the hope of depolarizing the populace in these countries to provide the impetus for national leaders to focus on global cooperation.

8. Conclusion

Building on work in neuroeconomics, we argue that a value-based framework may provide the most powerful understanding of the psychology and neuroscience of group cooperation. We believe that this multi-level approach provides a more comprehensive understanding of the mental and neural processes that underlie the decision to cooperate with others. However, we also believe that a more comprehensive model of cooperation requires an understanding of the role of groups and institutions in fostering trust and cooperation–even in the absence of value-based decision-making. Together, these frameworks provide a model for understanding the opportunities for fostering cooperation locally or at a global scale, which is necessary for addressing urgent crises like climate change.

References

Algan, Y., Cohen, D., Davoine, E., Foucault, M., & Stantcheva, S. (2021). Trust in scientists in times of pandemic: Panel evidence from 12 countries. *Proceedings of the National Academy of Sciences, 118*(40), Article e2108576118. https://doi.org/10.1073/pnas.2108576118

Andreoni, J. (1988). Why free ride? Strategies and learning in public goods experiments. *Journal of Public Economics, 37*, 291-304. https://doi.org/10.1016/0047-2727(88)90043-6

Armel, K. C., Beaumel, A., & Rangel, A. (2008). Biasing simple choices by manipulating relative visual attention. *Judgment and Decision Making*, *3*(5), 396–403.

Bakkour, A., Palombo, D. J., Zylberberg, A., Kang, Y. H., Reid, A., Verfaellie, M., Shadlen, M. N., & Shohamy, D. (2019). The hippocampus supports deliberation during value-based decisions. *ELife*, *8*, Article e46080. https://doi.org/10.7554/eLife.46080

Bartra, O., McGuire, J. T., & Kable, J. W. (2013). The valuation system: a coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *NeuroImage*, *76*, 412-427. https://doi.org/10.1016/j.neuroimage.2013.02.063

Bernhard, H., Fehr, E., & Fischbacher, U. (2006). Group affiliation and altruistic norm enforcement. *American Economic Review*, *96*(2), 217-221. https://doi.org/10.1257/000282806777212594

Bicchieri, C. (2002). Covenants without swords: Group identity, norms, and communication in social dilemmas. *Rationality and Society*, *14*(2), 192-228. https://doi.org/10.1177%2F1043463102014002003

Bicchieri, C. (2005a). *The grammar of society: The nature and dynamics of social norms*. Cambridge University Press.

Bicchieri, C. (2005b). Cooperation and communication: Group identity or social norms? In N. Gold (Ed.), *Teamwork: Multi-disciplinary perspectives* (pp. 102–135). Palgrave Macmillan UK. https://doi.org/10.1057/9780230523203_6

Bouwmeester, S., Verkoeijen, P. P. J. L., Aczel, B., Barbosa, F., Bègue, L., Brañas-Garza, P., Chmura, T. G. H., Cornelissen, G., Døssing, F. S., Espín, A. M., Evans, A. M., Ferreira-Santos, F., Fiedler, S., Flegr, J., Ghaffari, M., Glöckner, A., Goeschl, T., Guo, L., Hauser, O. P., … Wollbrant, C. E. (2017). Registered replication report: Rand, Greene, and Nowak (2012). *Perspectives on Psychological Science, 12*(3), 527–542. https://doi.org/10.1177/1745691617693624

Bowles, S., & Gintis, H. (2011). *A cooperative species*. Princeton University Press.

Brewer, M. B., & Kramer, R. M. (1986). Choice behavior in social dilemmas: Effects of social identity, group size, and decision framing. *Journal of Personality and Social Psychology, 50*(3), 543. https://doi.org/10.1037/0022-3514.50.3.543

Brzezinski, A., Kecht, V., Van Dijcke, D., & Wright, A. L. (2021). Science skepticism reduced compliance with COVID-19 shelter-in-place policies in the United States. *Nature Human Behaviour, 5*(11), 1519-1527. https://doi.org/10.1038/s41562-021-01227-0

Buchan, N. R., Brewer, M. B., Grimalda, G., Wilson, R. K., Fatas, E., & Foddy, M. (2011). Global social identity and global cooperation. *Psychological Science, 22*(6), 821-828. https://doi.org/10.1177%2F0956797611409590

Buckholtz, J. W., & Marois, R. (2012). The roots of modern justice: cognitive and neural foundations of social norms and their enforcement. *Nature Neuroscience*, *15*(5), 655–661. https://doi.org/10.1038/nn.3087

Camerer, C. F., & Fehr, E. (2004). Measuring social norms and preferences using experimental games: A guide for social scientists. In J. Henrich, R. Boyd, S. Bowles, C. Camerer, E. Fehr, & H. Ginti (Eds.), *Foundations of human sociality: Economic experiments and ethnographic evidence from fifteen small-scale societies* (pp. 55-95). Oxford University Press.

Carlson, R. W., & Crockett, M. J. (2018). The lateral prefrontal cortex and moral goal pursuit. *Current Opinion in Psychology, 24,* 77–82. https://doi.org/10.1016/j.copsyc.2018.09.007

Chang, L. J., & Sanfey, A. G. (2013). Great expectations: neural computations underlying the use of social norms in decision-making. *Social Cognitive and Affective Neuroscience, 8*(3), 277-284. https://doi.org/10.1093/scan/nsr094

Charness, G., Cobo-Reyes, R., & Jiménez, N. (2014). Identities, selection, and contributions in a public-goods game. Games and Economic Behavior, 87, 322-338. https://doi.org/10.1016/j.geb.2014.05.002

Charness, G., & Rabin, M. (2002). Understanding Social Preferences with Simple Tests. *The Quarterly Journal of Economics*, *117*(3), 817–869. https://doi.org/10.1162/003355302760193904

Chen, Y., & Li, S. X. (2009). Group identity and social preferences. American Economic Review, 99(1), 431-57. https://doi.org/10.1257/aer.99.1.431

Cialdini, R. B., Kallgren, C. A., & Reno, R. R. (1991). A focus theory of normative conduct: A theoretical refinement and reevaluation of the role of norms in human behavior. *Advances in Experimental Social Psychology*, *24*, 201–234. https://doi.org/10.1016/S0065-2601(08)60330-5

Curry, O. S. (2016). Morality as cooperation: A problem-centred approach. In T. K. Shackelford & R. D. Hansen (Eds.), *The evolution of morality* (pp. 27–51). Springer International Publishing AG. https://doi.org/10.1007/978-3-319-19671-8_2

Cutler, J., & Campbell-Meiklejohn, D. (2019). A comparative fMRI meta-analysis of altruistic and strategic decisions to give. *NeuroImage*, *184*, 227–241. https://doi.org/10.1016/j.neuroimage.2018.09.009

Dawes, R. M. (1980). Social dilemmas. *Annual Review of Psychology, 31*, 169-193. https://doi.org/10.1146/annurev.ps.31.020180.001125

Dawes, R. M., Van De Kragt, A. J. C., & Orbell, J. M. (1988). Not me or thee but we: The importance of group identity in eliciting cooperation in dilemma situations: Experimental manipulations. *Acta Psychologica, 68*(1), 83–97. https://doi.org/10.1016/0001-6918(88)90047-9

Declerck, C. H., & Boone, C. (2018). The neuroeconomics of cooperation. *Nature Human Behaviour, 2*(7), 438-440. https://doi.org/10.1038/s41562-018-0387-3

De Cremer, D., & Van Vugt, M. (1999). Social identification effects in social dilemmas: A transformation of motives. *European Journal of Social Psychology*, *29*(7), 871–

893. https://doi.org/10.1002/(SICI)1099-0992(199911)29:7<871::AID-EJSP962>3.0.CO;2-I

DeSteno, D. (2014). *The truth about trust: How it determines success in life, love, learning, and more*. Penguin.

DeWall, C. N., Baumeister, R. F., Gailliot, M. T., & Maner, J. K. (2008). Depletion makes the heart grow less helpful: Helping as a function of self-regulatory energy and genetic relatedness. *Personality and Social Psychology Bulletin, 34*(12), 1653–1662. https://doi.org/10.1177/0146167208323981

Dinesen, P. T. (2013). Where you come from or where you live? Examining the cultural and institutional explanation of generalized trust using migration as a natural experiment. *European Sociological Review, 29*(1), 114-128. https://doi.org/10.1093/esr/jcr044

Dinesen, P. T., & Sønderskov, K. M. (2018). Ethnic diversity and social trust. In E. Uslaner (Ed.), *The Oxford Handbook of Social and Political Trust* (pp. 175-204). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190274801.013.13

Doell, K. C., Pärnamets, P., Harris, E. A., Hackel, L. M., & Van Bavel, J. J. (2021). Understanding the effects of partisan identity on climate change. *Current Opinion in Behavioral Sciences, 42,* 54-59.

Domenech, P., Redouté, J., Koechlin, E., & Dreher, J. C. (2018). The neuro-computational architecture of value-based selection in the human brain. *Cerebral Cortex*, *28*(2), 585-601. https://doi.org/10.1093/cercor/bhw396

Dunbar, R. I. M. (1998). The social brain hypothesis. *Evolutionary Anthropology: Issues, News, and Reviews*, *6*(5), 178–190. https://doi.org/10.1002/(SICI)1520-6505(1998)6:5<178::AID-EVAN5>3.0.CO;2-8

Edney, J. J. (1980). The commons problem: Alternative perspectives. *American Psychologist, 35*(2), 131-150. https://doi.org/10.1037/0003-066X.35.2.131

Evans, A. M., Dillon, K. D., & Rand, D. G. (2015). Fast but not intuitive, slow but not reflective: Decision conflict drives reaction times in social dilemmas. *Journal of Experimental Psychology: General, 144*(5), 951–966. https://doi.org/10.1037/xge0000107

Fabbri, M. (2022). Institutional quality shapes cooperation with out-group strangers. *Evolution and Human Behavior, 43*(1), 53-70. https://doi.org/10.1016/j.evolhumbehav.2021.11.003

Falandays, J. B., & Spivey, M. J. (2020). Biasing moral decisions using eye movements: replication and simulation. In *Proceedings of the 42nd Annual Conference of the Cognitive Science Society*, 2553–2558.

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., & Fischbacher, U. (2004). Social norms and human cooperation. *Trends in cognitive sciences*, *8*(4), 185-190.

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

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. (2019). Viewing adaptive social choice through the lens of associative learning. *Perspectives on Psychological Science*, *14*(2), 175–196. https://doi.org/10.1177/1745691618792261

FeldmanHall, O., Montez, D. F., Phelps, E. A., Davachi, L., & Murty, V. P. (2021). Hippocampus guides adaptive learning during dynamic social interactions. *The Journal of Neuroscience, 41*(6), 1340–1348. https://doi.org/10.1523/JNEUROSCI.0873-20.2020

Fiedler, S., Glöckner, A., Nicklisch, A., & Dickert, S. (2013). Social value orientation and information search in social dilemmas: An eye-tracking analysis. *Organizational Behavior and Human Decision Processes*, *120*(2), 272–284. https://doi.org/10.1016/j.obhdp.2012.07.002

Fischbacher, U., Gächter, S., & Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics letters*, *71*(3), 397-404. https://doi.org/10.1016/S0165-1765(01)00394-9

Foddy, M., Platow, M. J., & Yamagishi, T. (2009). Group-based trust in strangers: The role of stereotypes and expectations. *Psychological Science, 20*(4), 419–422. https://doi.org/10.1111/j.1467-9280.2009.02312.x

Fosgaard, T., Jacobsen, C., & Street, C. (2021). The heterogeneous processes of cheating: Attention evidence from two eye tracking experiments. *Journal of Behavioral Decision Making*, *34*(1), 131–139. https://doi.org/10.1002/bdm.2200

Fowler, J. H., & Christakis, N. A. (2010). Cooperative behavior cascades in human social networks. *Proceedings of the National Academy of Sciences*, *107*(12), 5334–5338. https://doi.org/10.1073/pnas.0913149107

Gaesser, B., & Schacter, D. L. (2014). Episodic simulation and episodic memory can increase intentions to help others. *Proceedings of the National Academy of Sciences*, *111*(12), 4415–4420. https://doi.org/10.1073/pnas.1402461111

Gelfand, M. (2019). *Rule makers, rule breakers: Tight and loose cultures and the secret signals that direct our lives.* Simon and Schuster.

Ghaffari, M., & Fiedler, S. (2018). The power of attention: Using eye gaze to predict other-regarding and moral choices. *Psychological Science*, *29*(11), 1878–1889. https://doi.org/10.1177/0956797618799301

Gill, M. J., Packer, D. J., & Van Bavel, J. J. (2013). More to morality than mutualism: Consistent contributors exist and they can inspire costly generosity in others. *Behavioral and Brain Sciences*, *36*(1), 90. https://doi.org/10.1017/S0140525X12000799

Goette, L., Huffman, D., & Meier, S. (2006). The impact of group membership on cooperation and norm enforcement: Evidence using random assignment to real social groups. *American Economic Review*, *96*(2), 212-216. https://doi.org/10.1257/000282806777211658

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, *293*(5537), 2105-2108. https://doi.org/10.1126/science.1062872

Gu, X., Wang, X., Hula, A., Wang, S., Xu, S., Lohrenz, T. M., ... & Montague, P. R. (2015). Necessary, yet dissociable contributions of the insular and ventromedial prefrontal cortices to norm adaptation: computational and lesion evidence in humans. *The Journal of Neuroscience*, *35*(2), 467-473. https://doi.org/10.1523/JNEUROSCI.2906-14.2015

Gürerk, O., Irlenbusch, B., & Rockenbach, B. (2006). The competitive advantage of sanctioning institutions. *Science*, *312*(5770), 108-111. https://doi.org/10.1126/science.1123633

Hackel, L. M., Wills, & Van Bavel, J. J. (2020). Shifting prosocial intuitions: Neurocognitive evidence for a value based account of group-based cooperation. *Social Cognitive and Affective Neuroscience, 15*(4), 371–381. https://doi.org/10.1093/scan/nsaa055

Herrera, D., Novick, D., Jan, D., & Traum, D. (2011). Dialog behaviors across culture and group size. In C. Stephanidis (Ed.), *Universal access in human-computer interaction. Users diversity* (pp. 450–459). Springer. https://doi.org/10.1007/978-3-642-21663-3_48

Hare, T. A., Schultz, W., Camerer, C. F., O'Doherty, J. P., & Rangel, A. (2011). Transformation of stimulus value signals into motor commands during simple choice. *Proceedings of the National Academy of Sciences*, *108*(44), 18120-18125. https://doi.org/10.1073/pnas.1109322108

Hobbes, T. (2002). *Leviathan*. https://www.gutenberg.org/files/3207/3207-h/3207-h.htm (Original work published 1651)

Hodgson, G. M. (2006). What are institutions?. *Journal of Economic Issues, 40*(1), 1-25. https://doi.org/10.1080/00213624.2006.11506879

Hornsey, M. J., Harris, E. A., & Fielding, K. S. (2018). Relationships among conspiratorial beliefs, conservatism and climate scepticism across nations. *Nature Climate Change*, *8*(7), 614–620. https://doi.org/10.1038/s41558-018-0157-2

Hruschka, D. J., & Henrich, J. (2013). Institutions, parasites and the persistence of in-group preferences. *PLoS ONE, 8*(5), Article e63642. https://doi.org/10.1371/journal.pone.0063642

Hruschka, D., Efferson, C., Jiang, T., Falletta-Cowden, A., Sigurdsson, S., McNamara, R., Sands, M., Munira, S., Slingerland, E., & Henrich, J. (2014). Impartial institutions, pathogen stress and the expanding social network. *Human Nature, 25*(4), 567–579. https://doi.org/10.1007/s12110-014-9217-0

Hsu, M., Anen, C., & Quartz, S. R. (2008). The right and the good: Distributive justice and neural encoding of equity and efficiency. *Science, 320*(5879), 1092–1095. https://doi.org/10.1126/science.1153651

Hutcherson, C. A., Bushong, B., & Rangel, A. (2015). A neurocomputational model of altruistic choice and its implications. *Neuron*, *87*, 451-462. https://doi.org/10.1002/wcs.1571

Hutcherson, C. A., & Tusche, A. (2020). *Evidence accumulation, not "self-control," explains dorsolateral prefrontal activation during normative choice.* bioRxiv. https://doi.org/10.1101/2020.10.06.328476

Irwin, K. (2009). Prosocial behavior across cultures: The effects of institutional versus generalized trust. In S.R. Thye & E.J. Lawler (Eds.), *Altruism and prosocial behavior in groups.* Emerald Group Publishing Limited.

Jacquet, J., Hagel, K., Hauert, C., Marotzke, J., Rohl, T., & Milinski, M. (2013). Intra- and intergenerational discounting in the climate game. *Nature Climate Change, 3,* 1025-1028.

Jiang, T., Potters, J., & Funaki, Y. (2016). Eye-tracking social preferences. *Journal of Behavioral Decision Making*, *29*(2–3), 157–168. https://doi.org/10.1002/bdm.1899

Kallgren, C. A., Reno, R. R., & Cialdini, R. B. (2000). A focus theory of normative conduct: When norms do and do not affect behavior. *Personality and social psychology bulletin*, *26*(8), 1002-1012. https://doi.org/10.1177%2F01461672002610009

Khadjavi, M., & Lange, A. (2013). Prisoners and their dilemma. *Journal of Economic Behavior & Organization*, *92*, 163–175. https://doi.org/10.1016/j.jebo.2013.05.015

Kocher, M.G., Martinsson, P., Myrseth, K.O.R., & Wollbrant, C. E. (2012). *Strong, bold, and kind: Self-control and cooperation in social dilemmas* (ESMT Working Paper No. 12-01). https://dx.doi.org/10.2139/ssrn.1988286

Krajbich, I., Adolphs, R., Tranel, D., Denburg, N. L., & Camerer, C. F. (2009). Economic games quantify diminished sense of guilt in patients with damage to the prefrontal cortex. *Journal of Neuroscience*, *29*(7), 2188–2192. https://doi.org/10.1523/JNEUROSCI.5086-08.2009

Krajbich, I., Armel, C., & Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience*, *13*, 1292-1298. https://doi.org/10.1038/nn.2635

Krajbich, I., Bartling, B., Hare, T., & Fehr, E. (2015). Rethinking fast and slow based on a critique of reaction-time reverse inference. *Nature Communications*, *6*, 7455.

Krajbich, I., Hare, T., Bartling, B., Morishima, Y., & Fehr, E. (2015). A common mechanism underlying food choice and social decisions. *PLoS Computational Biology*, *11*(10), Article e1004371. https://doi.org/10.1371/journal.pcbi.1004371

Kramer, R. M., & Brewer, M. B. (1984). Effects of group identity on resource use in a simulated commons dilemma. *Journal of Personality and Social Psychology*, *46*(5), 1044-1057. https://doi.org/10.1037/0022-3514.46.5.1044

Kvarven, A., Strømland, E., Wollbrant, C., Andersson, D., Johannesson, M., Tinghög, G., ... & Myrseth, K. O. R. (2020). The intuitive cooperation hypothesis revisited: a meta-analytic examination of effect size and between-study heterogeneity. *Journal of the Economic Science Association*, *6*(1), 26-42.

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

Lin, S. Y., & Packer, D. J. (2017). Dynamic tuning of evaluations: Implicit racial attitudes are sensitive to incentives for intergroup cooperation. *Social Cognition, 35*(3), 245-272. http://doi.org/10.1521/soco.2017.35.3.245

Lindström, B., Haaker, J., & Olsson, A. (2018). A common neural network differentially mediates direct and social fear learning. *NeuroImage, 167,* 121-129. https://doi.org/10.1016/j.neuroimage.2017.11.039

Lim, S. L., O'Doherty, J. P., & Rangel, A. (2011). The decision value computations in the vmPFC and striatum use a relative value code that is guided by visual attention. *Journal of Neuroscience*, *31*, 13214-13223. https://doi.org/10.1523/jneurosci.1246-11.2011

Martinsson, P., Myrseth, K. O. R., & Wollbrant, C. E. (2012). Reconciling pro-social vs. selfish behavior: On the role of self-control. *Judgment and Decision Making, 7*(3), 304–315.

McLeish, K., & Oxoby, R. (2007). *Identity, cooperation, and punishment* (IZA Discussion Paper No. 2572). Institute of Labor Economics.

Mischkowski, D., & Glöckner, A. (2016). Spontaneous cooperation for prosocials, but not for proselfs: Social value orientation moderates spontaneous cooperation behavior. *Scientific Reports*, *6*, Article 21555. https://doi.org/10.1038/srep21555

Morishima, Y., Schunk, D., Bruhin, A., Ruff, C. C., & Fehr, E. (2012). Linking brain structure and activation in temporoparietal junction to explain the neurobiology of human altruism. *Neuron*, *75*(1), 73-79. https://doi.org/10.1016/j.neuron.2012.05.021

Nannestad, P., Svendsen, G. T., Dinesen, P. T., & Sønderskov, K. M. (2014). Do institutions or culture determine the level of social trust? The natural experiment of migration from non-western to western countries. *Journal of Ethnic and*

Migration Studies, *40*(4), 544-565. https://doi.org/10.1080/1369183X.2013.830499

Newell, B. R., & Le Pelley, M. E. (2018). Perceptual but not complex moral judgments can be biased by exploiting the dynamics of eye-gaze. *Journal of Experimental Psychology: General*, *147*(3), 409–417. https://doi.org/10.1037/xge0000386

Olsson, A., Knapska, E., & Lindström, B. (2020). The neural and computational systems of social learning. *Nature Reviews Neuroscience*, *21*(4), 197-212. https://doi.org/10.1038/s41583-020-0276-4

Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action*. Cambridge University Press.

Parkinson, C., Kleinbaum, A. M., & Wheatley, T. (2017). Spontaneous neural encoding of social network position. *Nature Human Behaviour*, *1*(5), 1-7. https://doi.org/10.1038/s41562-017-0072

Pärnamets, P., Balkenius, C. & Richardson, D. C. (2014). Modelling moral choice as a diffusion process dependent on visual fixations. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society,* 1132-1137. Cognitive Science Society. https://cogsci.mindmodeling.org/2014/papers/201/paper201.pdf

Pärnamets, P., Johansson, P., Balkenius, C., Hall, L., Spivey, M.J. & Richardson, D.C. (2015). Biasing moral choices by exploiting the dynamics of eye gaze. *Proceedings of the National Academy of Sciences*, *112*, 4170-4175. https://doi.org/10.1073/pnas.1415250112

Pärnamets, P., Shuster, A., Reinero, D. A., & Van Bavel, J. J. (2020). A value-based framework for understanding cooperation. *Current Directions in Psychological Science*, *29*(3), 227-234. https://doi.org/10.1177/0963721420906200

Pärnamets, P., Gill, M. J., Packer, D. J., & Van Bavel, J. J. (2021). *The role of visual attention in group cooperation* [Unpublished manuscript]. Department of Clinical Neuroscience, Karolinska Institutet.

Pärnamets, P., Wills, J. & Van Bavel, J. J. (2021). *Cultural differences in the automaticity of cooperation* [Unpublished manuscript]. Department of Clinical Neuroscience, Karolinska Institutet.

Pereda, M., Capraro, V., & Sánchez, A. (2019). Group size effects and critical mass in public goods games. *Scientific Reports*, *9*, Article 5503. https://doi.org/10.1038/s41598-019-41988-3

Platow, M. J., Foddy, M., Yamagishi, T., Lim, L. I., & Chow, A. (2011). Two experimental tests of trust in in-group strangers: The moderating role of common knowledge of group membership. *European Journal of Social Psychology*, *42*(1), 30-35. https://doi.org/10.1002/ejsp.852

Raihani, N. J., & McAuliffe, K. (2014). Dictator game giving: The importance of descriptive versus injunctive norms. *PLoS ONE*, *9*(12), Article e113826. https://doi.org/10.1371/journal.pone.0113826

Rand, D. G. (2016). Cooperation, fast and slow: Meta-analytic evidence for a theory of social heuristics and self-interested deliberation. *Psychological Science, 27*(9), 1192–1206. https://doi.org/10.1177%2F0956797616654455

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

Rand, D. G., Peysakhovich, A., Kraft-Todd, G. T., Newman, G. E., Wurzbacher, O., Nowak, M. A., & Greene, J. D. (2014). Social heuristics shape intuitive cooperation. *Nature Communications, 5*, 3677. https://doi.org/10.1038/ncomms4677

Reinero, D. A., Dikker, S., & Van Bavel, J. J. (2021). Inter-brain synchrony in teams predicts collective performance. *Social Cognitive and Affective Neuroscience, 16*(1-2), 43-57. https://doi.org/10.1093/scan/nsaa135

Robinson, E. J. H., & Barker, J. L. (2017). Inter-group cooperation in humans and other animals. *Biology Letters, 13*(3), Article 20160793. https://doi.org/10.1098/rsbl.2016.0793

Rothstein, B., & Stolle, D. (2008). The state and social capital: An institutional theory of generalized trust. *Comparative Politics, 40(*4), 441-459. https://doi.org/20434095

Rousseau, J.. (1998). *A discourse on a subject proposed by the Academy Of Dijon: What is the origin of inequality among men, and is it authorised by natural law?* (G. D. H. Cole, Trans.) https://constitution.org/2-Authors/jjr/ineq.htm (Original work published in 1754)

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

Satpute, A. B., & Lieberman, M. D. (2006). Integrating automatic and controlled processes into neurocognitive models of social cognition. *Brain Research*, *1079*(1), 86-97. https://doi.org/10.1016/j.brainres.2006.01.005

Schultz ,P. W., Nolan, J. M., Cialdini, R. B., Goldstein, N. J., & Griskevicius, V. (2007). The constructive, destructive, and reconstructive power of social norms. *Psychological Science, 18*, 429–434. https://doi.org/10.1111/j.1467-9280.2007.01917.x

Shadlen, M. N., & Kiani, R. (2013). Decision-making as a window on cognition. *Neuron*, *80*(3), 791-806. https://doi.org/10.1016/j.neuron.2013.10.047

Shadlen, M. N., & Shohamy, D. (2016). Decision making and sequential sampling from memory. *Neuron*, *90*(5), 927–939. https://doi.org/10.1016/j.neuron.2016.04.036

Shimojo, S., Simion, C., Shimojo, E., & Scheier, C. (2003). Gaze bias both reflects and influences preference. *Nature Neuroscience*, *6*(12), 1317–1322. https://doi.org/10.1038/nn1150

Smith, S. M., & Krajbich, I. (2018). Attention and choice across domains. *Journal of Experimental Psychology: General*, *147*(12), 1810. https://doi.org/10.1037/xge0000482

Smith, S. M., & Krajbich, I. (2019). Gaze amplifies value in decision making. *Psychological Science*, *30*(1), 116–128. https://doi.org/10.1177/0956797618810521

Solano, C. H., & Dunnam, M. (1985). Two's company: Self-disclosure and reciprocity in triads versus dyads. *Social Psychology Quarterly, 48*(2), 183–187. https://doi.org/10.2307/3033613

Sønderskov, K. M., & Dinesen, P. T. (2016). Trusting the state, trusting each other? The effect of institutional trust on social trust. *Political Behavior, 38*(1), 179-202. https://doi.org/10.1007/s11109-015-9322-8

Spadaro, G., Gangl, K., Van Prooijen, J. W., Van Lange, P. A., & Mosso, C. O. (2020). Enhancing feelings of security: How institutional trust promotes interpersonal trust. *PLoS ONE, 15*(9), Article e0237934. https://doi.org/10.1371/journal.pone.0237934

Spadaro, G., Graf, C., Jin, S., Arai, S., Inoue, Y., Lieberman, E., Rinderu, M. I., Yuan, M., Van Lissa, C. J., & Balliet, D. (2022). Cross-cultural variation in cooperation: A meta-analysis. *Journal of Personality and Social Psychology*. https://doi.org/10.1037/pspi0000389

Staats, A. W., & Staats, C. K. (1958). Attitudes established by classical conditioning. *The Journal of Abnormal and Social Psychology*, *57*(1), 37–40. https://doi.org/10.1037/h0042782

Stagnaro, M. N., Arechar, A. A., & Rand, D. G. (2017). From good institutions to generous citizens: Top-down incentives to cooperate promote subsequent prosociality but not norm enforcement. *Cognition, 167*, 212-254. https://doi.org/10.1016/j.cognition.2017.01.017

Stevens, J. R., & Hauser, M. D. (2004). Why be nice? Psychological constraints on the evolution of cooperation. *Trends in Cognitive Sciences*, *8*(2), 60–65. https://doi.org/10.1016/j.tics.2003.12.003

Strombach, T., Weber, B., Hangebrauk, Z., Kenning, P., Karipidis, I. I., Tobler, P. N., & Kalenscher, T. (2015). Social discounting involves modulation of neural value

signals by temporoparietal junction. *Proceedings of the National Academy of Sciences*, *112*(5), 1619–1624. https://doi.org/10.1073/pnas.1414715112

Sturgis, P., Brunton-Smith, I., & Jackson, J. (2021). Trust in science, social consensus and vaccine confidence. *Nature Human Behaviour, 5*(11), 1528-1534. https://doi.org/10.1038/s41562-021-01115-7

Szekely, A., Lipari, F., Antonioni, A., Paolucci, M., Sánchez, A., Tummolini, L., & Andrighetto, G. (2021). Evidence from a long-term experiment that collective risks change social norms and promote cooperation. *Nature Communications*, *12*(1), 1-7. https://doi.org/10.1038/s41467-021-25734-w

Teoh, Y. Y., Yao, Z., Cunningham, W. A., & Hutcherson, C. A. (2020). Attentional priorities drive effects of time pressure on altruistic choice. *Nature Communications*, *11*(1), 1-13. https://doi.org/10.1038/s41467-020-17326-x

Thielmann, I., Böhm, R., Ott, M., & Hilbig, B. E. (2021). Economic games: an introduction and guide for research. *Collabra: Psychology, 7(1),* 19004.

Thomas, A. W., Molter, F., Krajbich, I., Heekeren, H. R., & Mohr, P. N. C. (2019). Gaze bias differences capture individual choice behaviour. *Nature Human Behaviour*, *3*(6), 625–635. https://doi.org/10.1038/s41562-019-0584-8

Tinghög, G., Andersson, D., Bonn, C., Böttiger, H., Josephson, C., Lundgren, G., Västfjäll, D., Kirchler, M., & Johannesson, M. (2013). Intuition and cooperation reconsidered. *Nature*, *498*(7452), E1-E2. https://doi.org/10.1038/nature12194

Tusche, A., & Hutcherson, C. A. (2018). Cognitive regulation alters social and dietary choice by changing attribute representations in domain-general and domain-

specific brain circuits. *eLife*, *7*, Article e31185. https://doi.org/10.7554/eLife.31185

Van Bavel, J. J., & Packer, D. J. (2021). *The power of us: Harnessing our shared identities to improve performance, increase cooperation, and promote social harmony.* Little, Brown Spark.

van Dijk, E., & De Dreu, C. K. (2021). Experimental games and social decision making. *Annual Review of Psychology*, *72*, 415-438. https://doi.org/10.1146/annurev-psych-081420-110718

Van Lange, P. A. M., De Bruin, E. M. N., Otten, W., & Joireman, J. A. (1997). Development of prosocial, individualistic, and competitive orientations: Theory and preliminary evidence. *Journal of Personality and Social Psychology, 73*(4), 733–746. https://doi.org/10.1037/0022-3514.73.4.733

Van Lange, P. A.M., Joireman, J., Parks, C. D, & van Dijk, E. (2013). The psychology of social dilemmas: A review. *Organizational Behavior and Human Decision Processes, 120*(2)*,* 125-141. https://doi.org/10.1016/j.obhdp.2012.11.003

Van Lange, P. A. M., Liebrand, W. B. G., Messick, D. M., & Wilke, H. A. M. (1992). Social dilemmas: The state of the art. In W. B. G. Liebrand, D. M. Messick, & H. A. M. Wilke (Eds.), *Social dilemmas: Theoretical issues and research findings* (pp. 3-28). Pergamon Press.

Van Prooijen, J. W., Spadaro, G., & Wang, H. (2022). Suspicion of institutions: How distrust and conspiracy theories deteriorate social relationships. *Current Opinion in Psychology, 43*, 65-69. https://doi.org/10.1016/j.copsyc.2021.06.013

Verkoeijen, P. P., & Bouwmeester, S. (2014). Does intuition cause cooperation?. *PLoS ONE, 9*(5), Article e96654. https://doi.org/10.1371/journal.pone.0096654

Vikbladh, O. M., Meager, M. R., King, J., Blackmon, K., Devinsky, O., Shohamy, D., Burgess, N., & Daw, N. D. (2019). Hippocampal contributions to model-based planning and spatial memory. *Neuron*, *102*(3), 683-693. https://doi.org/10.1016/j.neuron.2019.02.014

Vlasceanu, M., Enz, K., & Coman, A. (2018). Cognition in a social context: A social-interactionist approach to emergent phenomena. *Current Directions in Psychological Science*, *27*(5), 369–377. https://doi.org/10.1177/0963721418769898

Wang, J., Suri, S., & Watts, D. J. (2012). Cooperation and assortativity with dynamic partner updating. *Proceedings of the National Academy of Sciences*, *109*(36), 14363–14368. https://doi.org/10.1073/pnas.1120867109

Weber, J. M., & Murnighan, J. K. (2008). Suckers or saviors? Consistent contributors in social dilemmas. *Journal of Personality and Social Psychology*, *95*(6), 1340-1353. https://doi.org/10.1037/a0012454

Wildschut, T., & Insko, C. A. (2007). Explanations of interindividual – intergroup discontinuity: A review of the evidence. *European Review of Social Psychology*, *18*(1), 175–211. https://doi.org/10.1080/10463280701676543

Wills, J., FeldmanHall, O., Meager, M. R., Van Bavel, J. J., Blackmon, K., Devinsky, O., Doyle, W. K., Luciano, D. J., Kuzniecky, R. I., Nadkarni, S. S., Vazquez, B., Najjar, S., Geller, E., Golfinos, J. G., Placantonakis, D. G., Friedman, D., Wisoff, J. H., & Samadani, U. (2018). Dissociable contributions of the prefrontal cortex in

group-based cooperation. *Social Cognitive and Affective Neuroscience, 13*(4), 349–356. https://doi.org/10.1093/scan/nsy023

Wilson, E. O., (1998). *Consilience: The unity of knowledge*. Alfred A. Knopf, Inc.

Wimmer, G. E., & Shohamy, D. (2012). Preference by association: How memory mechanisms in the hippocampus bias decisions. *Science*, *338*(6104), 270–273. https://doi.org/10.1126/science.1223252

Xiang, T., Lohrenz, T., & Montague, P. R. (2013). Computational substrates of norms and their violations during social exchange. *The Journal of Neuroscience*, *33*(3), 1099–1108. https://doi.org/10.1523/JNEUROSCI.1642-12.2013

Yamagishi, T., & Kiyonari, T. (2000). The group as the container of generalized reciprocity. *Social Psychology Quarterly*, *63*(2), 116-132. https://doi.org/2695887

Zhou, L., & Zhang, D. (2006). A comparison of deception behavior in dyad and triadic group decision making in synchronous computer-mediated communication. *Small Group Research*, *37*(2), 140–164. https://doi.org/10.1177/1046496405285125

Zucker, L. G. (1986). Production of trust: Institutional sources of economic structure, 1840–1920. *Research in Organizational Behavior*, *8*, 53–111.

Figure 1

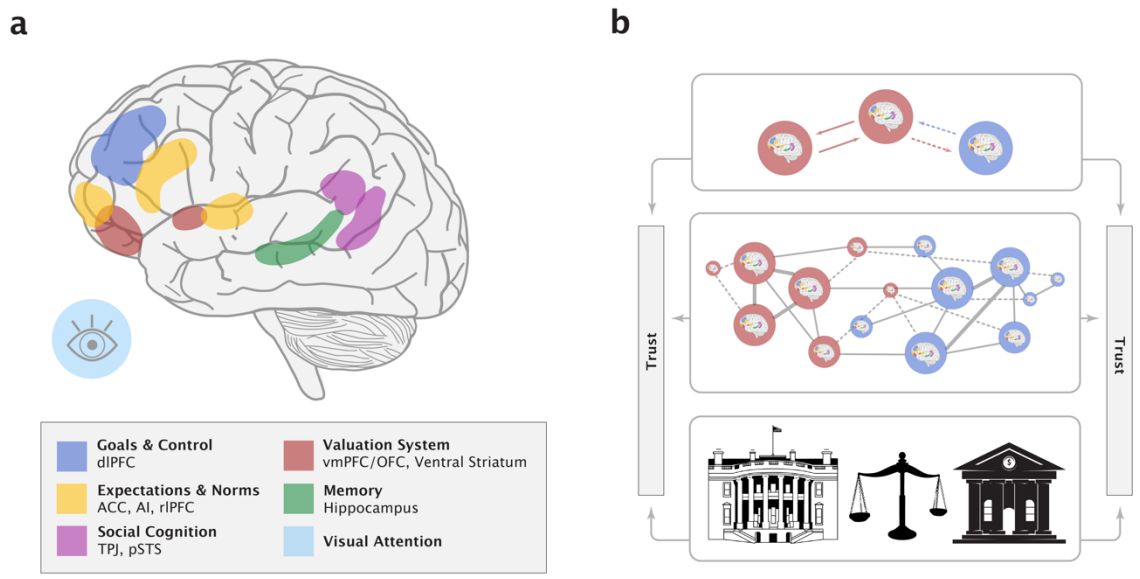

Figure 1. Summary of the value-based framework for cooperation (left) and a taxonomy showing that identities, social networks and institutions shape trust (right). The schematic of the value-based framework for cooperation (a) shows known modulatory inputs (denoted in the box) affect the computation of cooperative value. These inputs include goals and control, expectations and norms, social cognition, memory, and visual attention. These factors are implemented in various regions shown in the brain image and modulate the value system, which is implemented in the ventromedial prefrontal cortex (vmPFC) and ventral striatum. Decisions to cooperate are based, in part, on social trust (b), which is influenced by identities (in the top panel by the red vs. blue circles on the right), social networks (middle panel), and social institutions (bottom panel). dlPFC = dorsolateral prefrontal cortex; ACC = anterior cingulate cortex; AI = anterior insula; rlPFC = right lateral prefrontal cortex; TPJ = temporoparietal junction; pSTS = posterior superior temporal sulcus. (Adapted from Pärnamets et al., 2020 and created by Katie Brown, 2022).

Figure 2

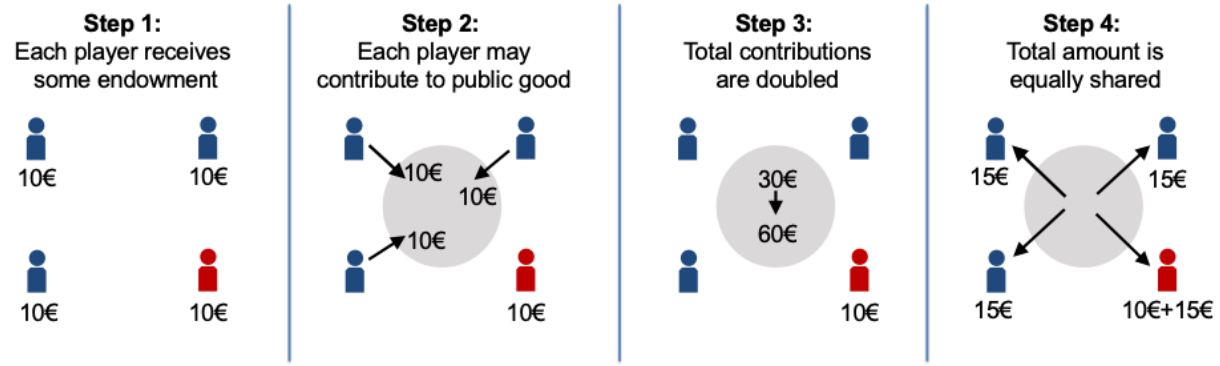

Figure 2. The Public Goods Game (PGG) involves four steps. In Step 1, each individual is given an endowment (usually a few dollars). In Step 2, individuals decide how much of their endowment they want to contribute to a shared public pool. In this example, the three blue players give all their endowment whereas the red player gives nothing (i.e., defect). In Step 3, all contributions to the public pool are increased (e.g., doubled). In the last step, all individuals receive an equal share from the public good. In this way, it is always in the collective best interest for everyone to cooperate, but always in their individual best interest to defect. Adapted from the Max-Planck-Institut (http://web.evolbio.mpg.de/evoltheo_corona/articles/AT_SocialDistDilemma/index_eng.html)

Figure 3

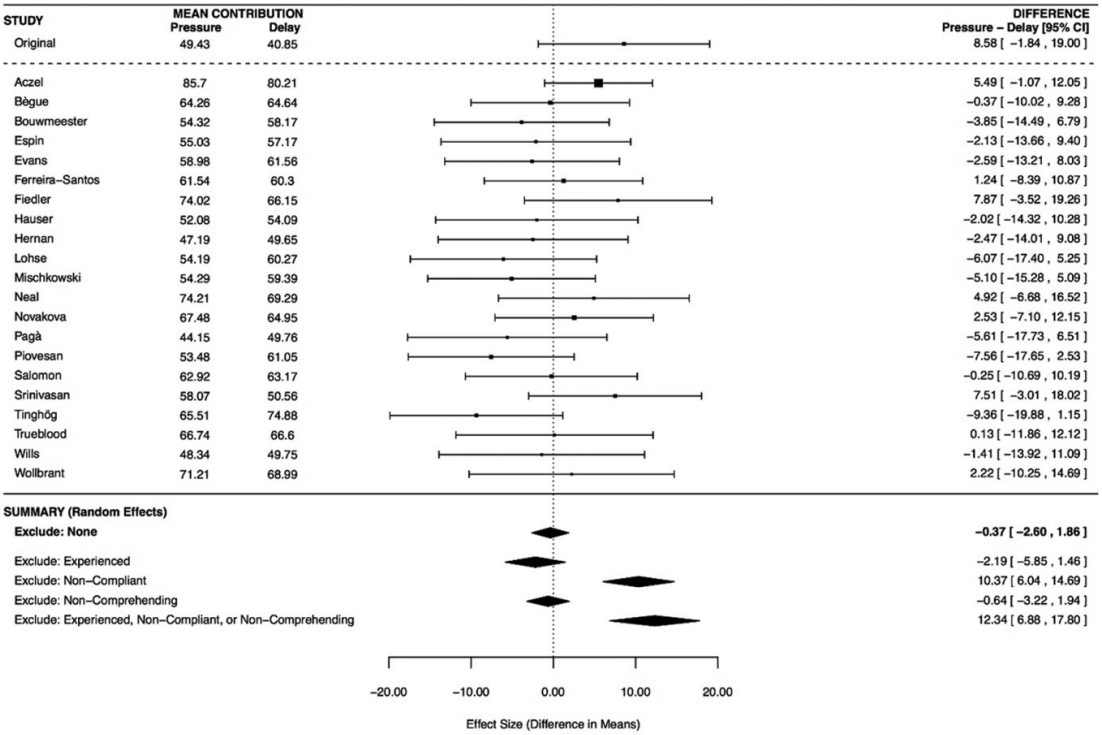

Figure 3. Forest plot and meta-analytic result for the difference in contributions between the time-pressure and forced-delay conditions from a large, multi-national registered replication of the intuitive cooperation hypothesis. Studies in the forest plot are listed alphabetically by the last name of the first author for that lab's study with the original result presented at the top. The mean difference for each lab is indicated by a square with the size corresponding to the inverse of the standard error of the difference score for that lab. The error bars indicate 95% confidence intervals around that laboratory's mean difference. The diamonds in the Summary section represent the results of random-effects meta-analyses of the RRR studies with the width representing a 95% confidence interval around the meta-analytic difference. The results provide mixed evidence for the intuitive cooperation hypothesis. The first diamond corresponds to the data in the forest plot and represents the primary planned meta-analysis with all participants (null effect). The next three diamonds show the meta-analytic difference after excluding experienced (null effect), non-compliant (significant effect), or non-comprehending (null effect) participants. The final diamond provides the meta-analytic difference when excluding participants who failed to meet any one of these criteria (significant

effect). Forest plots for the other meta-analyses are available at https://osf.io/scu2f/. (Figure adapted from Bouwmeester et al., 2017.)

Figure 4

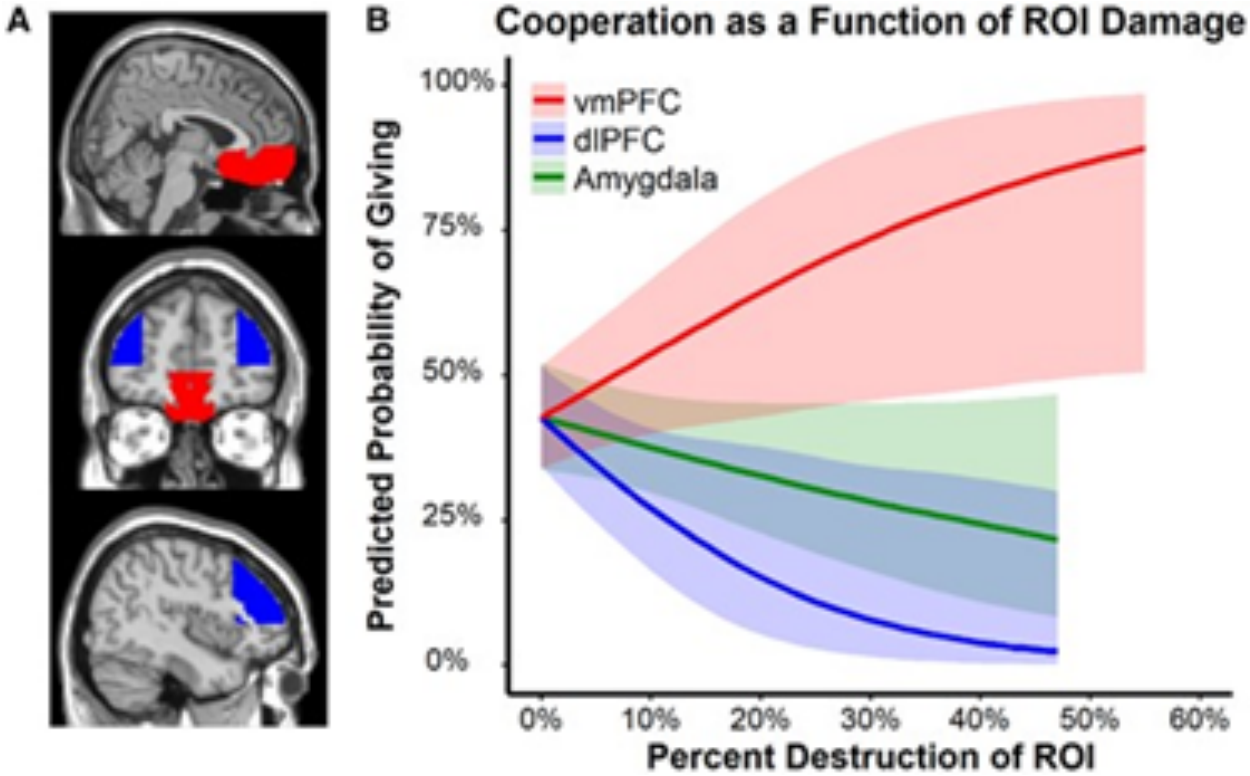

Figure 4. Cooperation as a function of damage to three brain regions. (A) vmPFC (red) and dlPFC (blue) regions of interest (ROIs) are pictured at the following slices: Sagittal X = 4 (top), Coronal Y = 40 (middle), Sagittal X = 44 (bottom). (B) Predicted probabilities of giving are plotted against damage within each ROI, after adjusting for external ROI damage. Lines are interpolated through the range of the observed levels of predictors. Bands indicate 95% confidence intervals. (Adapted from Wills et al., 2018).

Figure 5

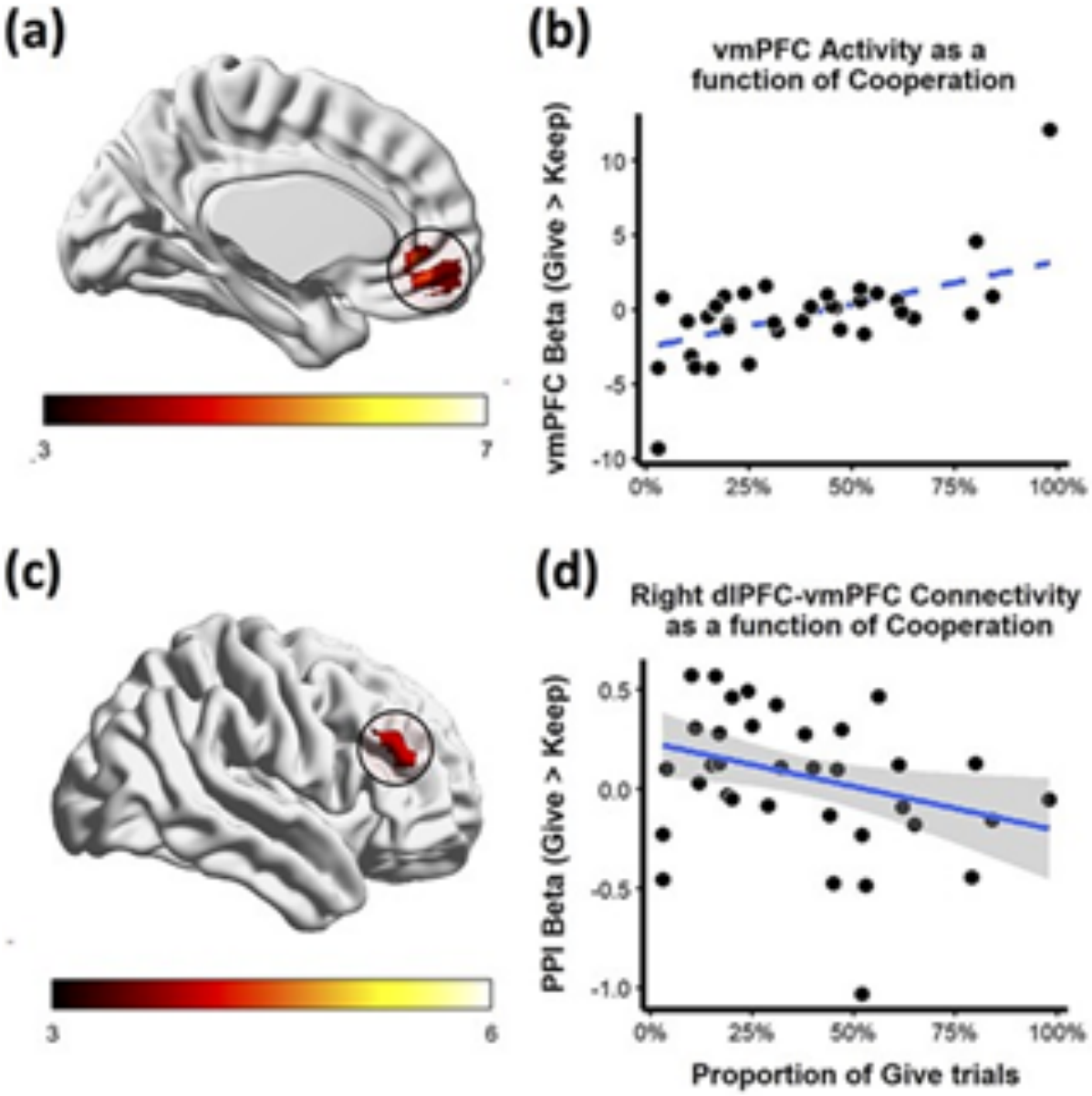

Figure 5. vmPFC activity and dlPFC-vmPFC connectivity is moderated by prosocial tendencies. Average cooperation moderates **(a)** BOLD response in vmPFC and **(c)** right dlPFC activity during Give (vs Keep) decisions. Color indicates magnitude of t statistic. As an alternate visualization, **(b)** vmPFC cluster betas (y-axis) for each participant (n = 35) are plotted against the proportion of cooperative trials (x-axis). **(d)** Right dlPFC-vmPFC PPI cluster betas (y-axis) are plotted against the proportion of cooperative trials (x-axis). Robust linear regression fits are displayed with blue lines and surrounding 95% confidence interval (Adapted from Hackel et al., 2020.)

Figure 6

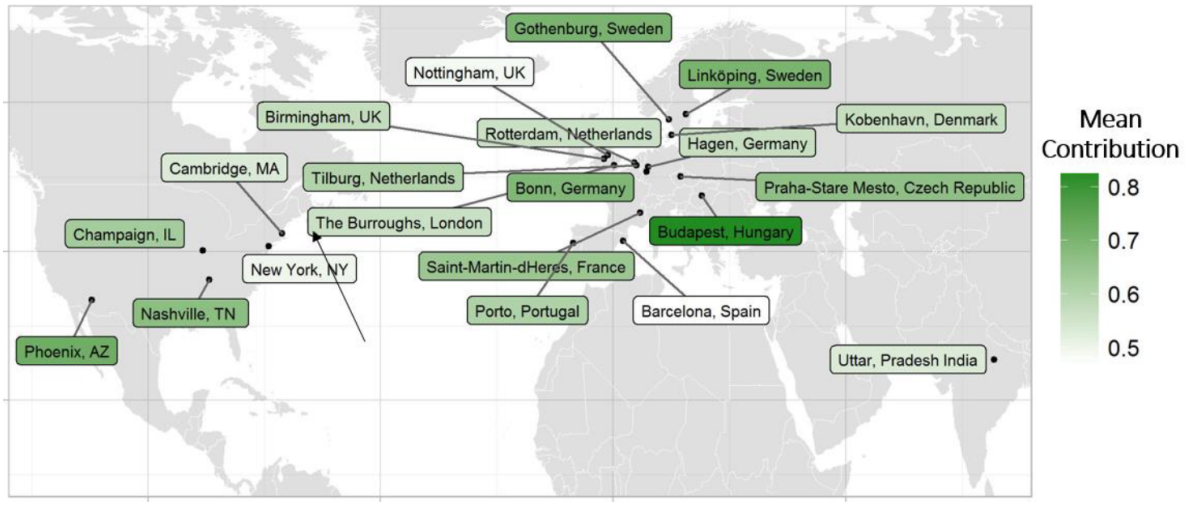

Figure 6. Variability in baseline cooperation. Each research site (k = 21) is represented by a colored bubble on the map. Sites where mean contributions were higher are represented in darker shades of green. For instance, the highest average contribution (80% of the total pot) was observed in Budapest, Hungary (Using data from Bouwmeester et al., 2017).