Hindawi
Journal of Theoretical Social Psychology
Volume 2022, Article ID 7259257, 19 pages
https://doi.org/10.1155/2022/7259257

Review Article

Beyond Reciprocity: Forgiveness, Generosity, and Punishment in Continuing Dyadic Interactions

Ann C. Rumble,[1] **Kevin Willcox**,[1] **Hirotaka Imada**,[2] **and Dejah Yansen**[1]

[1]*Department of Psychological Sciences, Northern Arizona University, Flagstaff, AZ 86001, USA*
[2]*School of Psychology, University of Kent, Canterbury, UK*

Correspondence should be addressed to Ann C. Rumble; ann.rumble@nau.edu

Received 7 March 2022; Accepted 3 May 2022; Published 20 June 2022

Academic Editor: Richard Crisp

There is a long-standing debate in philosophy and the social sciences about how selfishness and cooperation function in dyadic social exchanges. Dyads are the foundation of our social lives, and reciprocity has long been considered the dominant strategy for dyadic interactions. We will argue the repertoire of human behavior during social exchanges ranges from punishment to generosity, and that the nuances of the relationship and interaction will dictate which behavior is likely to occur. We will examine emotional consequences of punishment, reciprocity, and forgiveness in long-term dyadic social exchanges. Finally, we argue that dyads move beyond reciprocity to a more forgiving, generous strategy to reestablish cooperation, and continue the relationship when noncooperation arises, once the motivations shift has occurred.

1. Introduction

Beyond reciprocity: forgiveness, generosity, and punishment in continuing dyadic interactions:

"... the life of man solitary, poor, nasty, brutish, and short. The condition of man ...

is a condition of war of everyone against everyone," Thomas Hobbes.

"...No, it is better to be generous, and in the end more profitable, for it gains gratitude for us and love," Mark Twain [1].

There is a long-standing debate in philosophy and the social sciences about how selfishness and cooperation function in social exchanges. Thomas Hobbes professed that man is an inherently selfish creature, and behavioral economists have been operating off this assumption for decades [2, 3]. Social exchange research, however, shows that people often cooperate even when it does not seem to be their immediate material advantage [2]. These two schools of thought seem to contradict each other, but human tendencies toward selfishness and cooperation are not so simple as to be bipolar. Whether a person behaves selfishly or cooperates depends on a number of factors including the individual's general tendencies, group size, details of the group members, details of the situation, as well as potential benefits and costs [4–6]. This paper aims to provide a more thorough understanding of how various interaction dynamics affect cooperation and outcomes. We hope this initial review of the literature encourages behavioral economists and psychologists to further explore the complexities of dyadic relationships.

We first provide context by reviewing views of human nature as either inherently selfish or cooperative. Then, we examine how dyadic interactions (the foundations of social interaction) and larger group interactions have different dynamics. Third, we critically review interaction strategies from the behavioral economics game theory literature. In particular, we look at the material, social and emotional consequences of punishment, reciprocity, and forgiveness in long-term dyadic social exchanges. Finally, we investigate how humans change their interaction strategies through the course of a relationship (see figure one for our proposed model).

2. Selfishness and Cooperation

Thomas Hobbes [3] claimed that without strong governmental influence or control, humans would consistently behave selfishly. In game theory literature, selfishness has been defined as the desire to maximize one's own immediate outcomes without regard for others or interpersonal realities of extended interactions. Behavioral economists coined a term for the person that acts this way: *Homo economicus* [2, 7–11]. Homo economicus individuals follow a strategy of near-constant defection in the hopes of maximizing personal outcomes. In the context of the classic one-shot prisoner's dilemma game, in which two people have the option to either perform cooperatively or selfishly based on their partners' actions, theoretically the best strategy for maximizing individual outcomes is to defect (allowing the defector to keep all their own money and any money that their partner gives to them). In reality, only a small number of Hobbes-like individuals (7–8% of the population) act in this manner [11].

One major problem with choosing constant defection as a strategy is that it will likely inspire one's partner to defect also, either by leaving the interaction or punishing the defector, making it costly for the defector [12, 13]. Another major problem with constant defection is that if *everyone* acts this way, outcomes will be worse for everyone involved and threaten the long-term stability of the interaction [14, 15].

Contrary to Hobbes's assertion that all humans act like homo economicus individuals, social exchange research has demonstrated that most people do not always defect [5, 10, 11, 15, 16]. That is not to say that human behavior is saintly. While there are individuals who consistently cooperate, they constitute a small minority of the population. Both consistent defectors and consistent cooperators suffer decreased material outcomes in social exchanges [17–19].

The true state of human behavior appears to be that we are neither solitary devils nor altruistic angels. Rather, research has determined in multiple settings that defection, reciprocity, cooperation, generosity, and altruism are all part of the human behavioral repertoire [6, 10, 20–23]. We are not simple creatures limited to behaving in one way; we use myriad behavioral strategies influenced by personal and social preferences, circumstances, and external constraints [24–26].

The purely selfish homo economicus is an unsubstantiated reductionist theory of human behavior, with immediate self-interest as the primary goal and motivation. It does not capture the complex nature of human psychology or motivation [27]. It limits human dyadic relationships. It discounts motives for acting contrary to personal material outcomes, disregards how decision-making changes over time, and does not consider the negative impact of selfishness on dyadic membership. Psychology has long recognized that a purely economic explanation for human behavior is inadequate [4, 27–30]. The fields of economics and psychology need a more subtle understanding of dyadic behavior that accounts for the complexities of human interaction, such as how individual behavior is guided by impression formation, relationship maintenance, emotions, outcomes, and the decision-making processes. The homo economicus view of decision-making overlooks the dynamic, lasting, and complex nature of dyadic interactions [27, 31].

3. Selfishness and Cooperation in Dyads and Groups

Selfishness and cooperation function differently depending on the size of the group. However, many behavioral economists and psychologists do not yet acknowledge the distinction and subsequently lack a subtle understanding of dyadic interactions [32]. These differences reveal important information about long-term human interactions. We investigate dyads in this paper not only because of their importance to humans in daily life, but to emphasize how various dynamics need to be studied more thoroughly. We hope that by analyzing one type of relationship (long-term dyadic interactions), we will show the benefits of understanding the subtle dynamics of specific types of relationships. To do this, we will first talk about the importance of the dyad as a social unit. We will then investigate how dyads differ from larger groups and how they should therefore be considered and studied separately.

4. The Dyad as a Critical Social Unit

Hobbes's description of the human condition as *solitary* is a fundamental flaw in his argument, as humans are inherently social creatures who have developed numerous mechanisms for maintaining necessary connections with others [4, 27, 33–35]. Caporael and colleagues (2006 with [36]) argue that, in order to survive hostile environments, our human ancestors had to form relationships and groups with kin and nonrelatives alike to perform coordinated tasks, such as food gathering and defense. Furthermore, what Dawkins calls the *selfish gene,* [7] provides an inadequate explanation for coordinated human behaviors because it fails to account for the social and psychological motivations that guide human behavior [4, 27]. Social groups cannot be reduced to an aggregate, nor can behavior within a group be explained solely based on the individual members' self-interest [4, 27, 37]. What was necessary for human evolution in a hostile environment is *obligated interdependence* [4, 21, 37, 38], in which the behaviors that sustain the group are critical for survival, and purely self-interested behavior is potentially detrimental to basic survival. Sociality, more than solitary self-interest, is necessary, as it fosters cooperation and coordination essential for maximizing performance on survival tasks [4, 22, 33, 34, 37–40].

Obligatory interdependence is cultivated within a variety of groups, and each group develops specialized skills to complete group-level tasks. Caporael [4] postulates there are four "core configurations" of groups, differentiated by group size and task: the macrodeme, the deme, the task group, and the dyad. Macrodemes are large groups ($n < 300$) necessary for the spread of language and knowledge. Demes ($n \approx 30$) form our social identity and construct our shared reality or

culture. Face-to-face or task groups help with coordinated actions (such as hunting large game) that require specialized skills, such as shared cognition, memory, and perception. Dyads comprise two interacting partners whose primary evolutionary purposes include reproduction, childcare, and tasks related to basic survival, which help to maintain the fitness of the larger group [4]. Dyads are a critical social unit with different dynamics than larger groups and should therefore be researched separately.

5. Dynamics of Dyads vs. Larger Groups

Modern dyadic relationships form the foundation of our social interactions: parent/child, spouses, business partners, boss/employee, pairs of friends, etc. The meaning of the familiar idiom "two's company, three's a crowd" emphasizes how dyads are transformed when reassembled into larger groups. Research focusing on the relationship between group size and cooperation found that "three is a crowd" when trying to maintain cooperation across time in a noisy environment; the likelihood of cooperating after defection was found to decrease for groups involving 3–5 individuals when compared to dyads [41]. Furthermore, economics research finds that groups larger than two are less likely to cooperate [42–44].

Dyads differ from larger groups in cooperation (and other behaviors) because the dynamics that exist in a dyad are fundamentally different. In larger groups, individuals' choices and behaviors may be masked by any number of group-specific features. For example, in the provision of a public good like National Public Radio, all group members can consume the resource, but not all contribute to the resource [28, 45]. A noncontributor's choice can be masked by uncertainty, group size, and anonymity [46–48]. Compared to larger groups, dyads are *unique* in that all interaction is direct and lacks any ambiguity in regard to the source of a behavior or action. However, that is not to say there is complete transparency in dyads; ambiguity related to motives, expectations, and circumstances may be present [49]. Thus, in order to maintain cooperative dyadic relationships long term, partners remain attuned to both their own and the other's behavior, while also seeking an understanding of what motivates a particular behavior and how it influences the pair's future behavior [26, 49–54].

It should be noted there is a range of dyadic relationships, business partners, friends, and romantic partners. While the context of these relationships is quite different and influence by cultured in important ways, we rely on understanding dyads from an interdependence theory perspective. Specifically, we are utilizing an interdependent analysis to understand the *structure* of the interactions. Interdependence theory has been used to model interactions between romantic partners [50, 54–58] and dyadic interactions, in general [26, 53, 59].

Rusbult and Kelley [59] have argued that interdependence theory can be used to model moment-by-moment decision-making in dyadic relationships, thus illustrating how and why behaviors develop and continue and the forces that shape them [26, 50, 53, 54, 57, 59, 60]. Kelley and colleagues [59] theorize that behavior in a dyadic relationship result from each individual's expectations of a situation and the actual situation in which the interaction occurs. Initially, individuals find themselves in a given situation (e.g., a negotiation) that places its own constraints on behavior. Then, individuals can transform the given situation into an effective one through their own motives, emotions, perceptions, and habits. For example, an individual can engender trust to transform a zero-sum situation (one gains, the other loses) into a nonzero-sum interaction in which both can gain. If the partners have high trust, they may be more willing to cooperate or compromise due to faith in the others' intentions [61–63].

In order to fully understand dyadic relationships, researchers must be able to understand how interactions transform over time. We must consider how past interactions led to the present situation, and choices in the moment lead to different situations in the future [56, 58, 60, 64, 65]. Homo economicus approaches every situation the same (as an opportunity to procure resources immediately). For most people, however, a given situation in an ongoing dyadic interaction does not remain static but shifts due to previous choices and decisions made by each partner. For example, a person can start an interaction with reciprocal behavior, become more generous due to some mutually trusting behavior, but then get angry and punish the other person because they became uncooperative. It is crucial to note these transformations after the initial interaction between the members, as economists traditionally assumed a relatively static psycho-behavioral strategy (homo economicus or strong reciprocity) [66–68].

6. Specifics of Dyadic Interactions

Dyads are broadly responsible for tasks that require coordinated action or the formation of exchange relationships [26, 59]. During a coordination task, the two individuals must coordinate their actions to achieve the desired outcome. To give an example, say a spousal pair wants to clean their home. If both individuals decide to wash the dishes, then the laundry is left undone. If they coordinate their actions, one can do the dishes while the other does laundry. Alternatively, the dyad may face an exchange task, which may be considered similar to a negotiation, as each individual has some resource (tangible or intangible) to exchange with the other [26, 54, 59].

The difficulty that dyads face is that during either of these tasks, coordination or exchange, the members may either be in agreement or disagreement about the possible outcome(s) [26, 54, 59]. In interactions where each partner desires the same outcome, conflict is less likely, and each partner will be satisfied. However, if each partner has contrasting wants and/or needs, a compromise is more difficult to establish, resulting in contentious interaction and reduced overall satisfaction. Returning to the previous example of the spousal dyad, suppose washing dishes is perceived to be a relatively easy household task, which may explain why both individuals in the dyad wish to complete this task over the less desirable laundry one. If the dyad is not in agreement,

this will influence future interactions [26, 54, 59]. If in the first week one partner does the less desirable task of laundry, they will likely have to switch tasks the following week in order to remain satisfied [57, 69]. The situation and choices available to each individual are greatly shaped by what has happened previously and how motivated they are to remain in the interaction [26, 53, 54, 59, 70].

At the onset of any dyadic interaction, whether it takes place in the business world or in personal relationships, individuals struggle to establish a pattern of behavior that will encourage cooperation, reduce feeling disadvantaged, and maximize outcomes, while simultaneously promoting continuation of the interaction; if they do not succeed, the relationship is unlikely to continue [71]. Each individual monitors interactions to determine if needs are being met and how the other individual makes choices [26, 53, 54, 59]. During these initial interactions, partners may sometimes disagree about the outcome and face a choice between self-interest and relationship-centered interest. Holmes and Rempel [72] termed these early interactions *diagnostic* since they provide both partners with information about the other's motives and greatly shape future interactions and the overall relationship. Suppose a couple has met through an online dating site and are trying to decide how to spend their first date. Chris wants to attend a concert, while Tom suggests a football game; if Tom is willing to concede and go to a concert, it demonstrates a willingness to forgo his own self-interest for Chris and the relationship [10, 73]. It signals to Chris that Tom is not just interested in his own gains and interest but considers Chris and the relationship in making decisions and increases Chris's level of satisfaction [73]. However, at this early stage of the relationship, Tom is not committed enough to forgo his self-interest continually and risk feeling disadvantaged, and he will end the relationship unless during the next diagnostic situation Chris is willing to compromise [69, 74–77]. Thus, during the beginning stages of an interaction, reciprocity of behaviors/choices may be the norm. When both partners employ the rule of reciprocity at the initial stage, they encourage cooperation and increase the chance of an ongoing relationship.

7. Reciprocity and Its Achilles' Heel

Reciprocity is responding in kind to the action of an interaction partner, resisting noncooperation, and making amends for harm done [78]. Reciprocity may have developed over time as a specialized capability for dyads because it can enhance and maintain the fitness of the dyad, and thus the entire group. If a dyad develops a rule for behavior based on reciprocity, this can have a long-term stabilizing effect on the interaction and help dyads negotiate potentially risky early interactions [72, 79].

The rule of reciprocity pervades human interactions [78]. Research on topics as varied as attitude change, marketing, economic behavior, political behavior, self-disclosure, and close relationships have all demonstrated that individuals in dyads follow the rule of reciprocity [78, 80–82]. The persuasion literature, for example, has demonstrated that an individual is more persuaded by someone who earlier conceded to the individual's persuasive argument than by someone who did not concede [81]. Nonhuman primates, our closest animal kin, also follow the rule of reciprocity with regard to food sharing [83].

Much of the work on reciprocity comes from the experimental game theory literature, specifically two-person mixed-motive games, such as the Prisoner's dilemma (PDG) or the Trust game [71, 84]. While some of the work on reciprocity is based on simulations (e.g., [71, 85–87]), which has limited external validity, our analysis relies on an understanding of the behavioral research (e.g., [15, 29, 88–92]). The prisoner's dilemma and its variants were designed to understand behavior in a setting where immediate personal interests may be counter to long-term personal interests and/or the best interests of the dyad [15]. Axelrod [71] devised two computer tournaments in order to understand which behavioral strategy was most effective in maximizing outcomes during the classic PDG. Axelrod [71] designed a multitrial simulated PDG and solicited strategies from experimental game experts across a number of fields. Then in a round-robin format, each strategy played the 14 other strategies, plus itself, and a completely random strategy. In the first round of the tournament, tit for tat (TFT), or strict reciprocity, was the simplest strategy and won the most points on average during a 200-trial PDG. In the second round of the tournament, Axelrod [71] put out a wider call for strategies and provided the results from the first round of the tournament to interested parties prior to the second tournament. Sixty-two strategies were submitted and played a multitrial PDG (number of trials was determined probabilistically in order to avoid end game effects), in the same round-robin format. Again, TFT was the clear winner with the highest average score.

Compared to other strategies in the tournament, TFT maximizes outcomes mainly because it is nice, forgiving, clear, and retaliatory [71]. TFT begins each interaction with a cooperative act, thus it is "nice." It is easy to discern, thus the pattern is "clear." It is "forgiving" because it reacts to each event individually, so past uncooperative behavior is forgiven. But if the other does not cooperate, TFT will retaliate with noncooperation. These characteristics of TFT work to maximize outcomes and ensure the social exchange will continue, by increasing the benefit of cooperative behavior and reducing benefits of chronic defection [88, 90, 93]. Unlike the other strategies, TFT has established that it could not be taken advantage of (retaliation), while at the same time signaling its willingness for mutual cooperation (nice and forgiving). More selfish and less nice strategies, as well as less selfish strategies, were less successful in the tournament at maximizing outcomes and promoting mutual cooperation.

Strict reciprocity is defined as "nice" because it begins an interaction with cooperation, and it is never the first to defect. During Axelrod's tournaments [71], being nice was one characteristic the top eight strategies shared in the first round of the tournament. Being nice or cooperative at the beginning of an interaction is important for impression formation and increasing chances of mutual cooperative actions over time. An initial act of cooperation reduces the

amount of time individuals will take to reach 100% cooperation during a social exchange [94]. Also, choosing cooperation at the outset increases the probability of mutual cooperation and the probability that cooperation will last throughout the interaction [88, 90]. If individuals choose to defect at the onset of an interaction and then switch to cooperation, they are less likely to elicit continued cooperation than those who cooperate at the onset. If the first act is defection and is not followed by cooperation, this further reduces the chances of continued cooperation [88, 90, 93]. Lount and colleagues [95] found that defection during the first two trials of a PDG had a far greater negative impact on cooperation than defection on trials 6 & 7, or 11 & 12 out of 30 trials. Niceness serves to signal a willingness to be cooperative that influences the other's impression of the individual as cooperative and trustworthy [88].

Further evidence for the power of niceness at the onset of an interaction can be found in research on trust in *zero acquaintance* interactions [29]. Dunning and colleagues [29] examine trust as a basic *moral* norm of behavior that leads to initial cooperative behavior during the trust game between *zero acquaintances,* individuals who will never interact or meet, and remain completely anonymous. Dunning and colleagues [29] state that a moral norm is an internalized norm that does not require enforcement and demonstrates respect for the other person's moral character. During their experiment, first responders in the trust game could choose to keep or give their money ($5) to the second responder, who, after the money was increased ($20), could keep or split the money with the first responder. Homo economicus choice models dictate keeping the money during the first decision; however, most people did not keep the money, but gave it to the second responder. They made the cooperative, trusting choice initially in both the PDG and the Trust game. This not only provided the foundations for partner cooperation, but also allowed for potential relationship building long term. A first responder did not want to communicate she/he thought the second responder was untrustworthy or lacked moral character and believed that keeping the money would have implied this about the second responder. Dunning and colleagues [29] argue that once the cooperative/trusting choices are made, it can provide what Holmes and Rempel [72] term "reciprocal reassurance," and the interaction has the potential to begin and to remain cooperative, but without that initial trusting choice the interaction may never even begin [72].

Responding in kind to cooperation, even after a period of noncooperation, is equally as important as being nice. Axelrod [71] recommends that individuals should forgo envy by not worrying about others' outcomes. They should reciprocate cooperation with cooperation, regardless of the previous noncooperative act/s. Envy can lead to negative emotions that drive competitive behavior rather than cooperative behavior, which will challenge the long-term success of any human relationship [96, 97]. TFT is more likely than less forgiving strategies to maximize outcomes and to increase cooperation over time [71]. The immediate reciprocation of cooperation is important in changing the long-term outcomes of an interaction, and for encouraging future cooperation [90].

While cooperation promotes continued cooperation, unconditional cooperation can be detrimental when exploitation is possible. TFT reduces the odds of exploitation; unconditional cooperation is likely to promote exploitation and the experience of being a sucker [17–19, 71]. One of the reasons why strict reciprocity is successful is that it *does* respond to defection with defection. Humans over time may have adopted the reciprocity rule in part because it reduces the chance of being exploited, thus avoiding the negative emotional consequences of "feeling duped" [98].

Feeling like a sucker happens when a person feels they were treated less than equitably in an interaction. Participants perceive the situation as inequitable because their partner violated the rules of fairness governing the situation. This leads to a perception that the sucker has been taken advantage of, partly as a result of the sucker's own decisions [98]. Effron and Miller [99] found that individuals avoid making decisions when they could have their trust violated. Individuals also experience more self-blame if their trust was violated than when their trust was well placed. Even so, individuals did not avoid all risky decisions, nor did they become more risk averse due to a loss for other reasons, only trust violations led them to be particularly risk averse. In situations when an individual is duped by another, it can lead them to have less motivation to cooperate. It can also cause the sucker to perceive that their trust has been violated. Individuals are highly motivated to detect chronic free riders [100], and thus avoid the negative emotional consequences of feeling duped, which include frustration, anger, and self-blame.

While strict reciprocity is nice at the beginning of an interaction and forgiving during an interaction, this does not compromise its ability to react to noncooperative acts. By responding in kind to acts of noncooperation, strict reciprocity also reduces chronic noncooperation over the course of an interaction [90]. Tit for Tat will remain effective at maximizing outcomes and fostering ongoing interactions, *if* cooperation can be maintained, which, we will see shortly, is an important caveat in real world interactions.

8. The Downside of TFT

The Achilles' heel of TFT is a cycle of mutual defection resulting from immediate retaliation, which reduces outcomes and the chance for mutual cooperation in long-term interactions. Since Axelrod's tournaments, numerous researchers have pointed out that TFT, i.e., strict reciprocity, suffers a potentially fatal flaw in long-term interactions [85–87, 92, 101]. TFT is a *reactive* strategy that makes it impossible to change the course of a noncooperative exchange, which leads ultimately to lower outcomes for all [85–87, 92, 101]. If one of the interaction partners decides to test the other or the constraints of the situation by defecting, a strictly reciprocal partner will retaliate, which may lead the first partner to respond in kind. If both continue to use TFT and reciprocate defection, and neither changes their

behavior by taking a risk and cooperating, the interaction will continue with mutual defection until it ends.

The immediate reciprocation of retaliation can then present a particular challenge to ongoing dyadic interactions when the interaction or relationship is as or more important than the immediate outcomes available. A cycle of retaliation may reduce satisfaction in a dyadic interaction, reduce overall outcomes, and lead one or both partners to exit the interaction. If we are motivated to maintain a particular relationship long term, strict TFT limits our ability to foster and maintain long-term cooperation.

Because chronic defection can frustrate cooperation, it is important to understand why people defect in the first place. Defection may result from multiple factors, including intentional defection, situational uncertainty, personal ambiguity in continuous social exchanges, and incomplete information. Van Lange and colleagues [92] have demonstrated that when noncooperation is present, whether intentional or unintentional, cooperation will decrease. During a dyadic extended PDG, participants cooperated less over time when noncooperation was present, entering a defection loop [102].

Almost all social exchanges will, at some point, contain intentional and/or unintentional noncooperation. Individuals are not always able to discern the reason behind an incident of noncooperation and may create attributions for the behavior based on limited information that may reduce the chance for cooperation [86, 103]. If both interaction partners follow a strict reciprocal strategy, this can have a detrimental effect on the outcomes and reduce their satisfaction with the interaction [85–87].

8.1. Intentional Noncooperation. Intentional noncooperation can arise from a number of sources, including selfish motives and concern for being exploited. Hobbes and some behavioral economists [3, 7] contend that most human interactions are guided by selfish motives to increase immediate material outcomes without regard for the other or environmental constraints. In some situations, self-interested motives do override the concern for others, and individuals do act in a selfish manner, such as in competitive markets [104]. However, some individuals enter dyadic relationships with more pro-social motives, which foster more cooperative interactions [6, 23, 105]. Social value orientation (SVO) describes how individuals vary on the amount of concern they have for their own outcomes, their partner's outcomes, and the equality between themselves and their partner [6, 23, 105, 106].

SVO specifies four types of individuals (though in practice studies often measure only 2 or 3 types). The "pro-social individual" is of two types. One, often referred to as "pro-social" or "cooperative," is most concerned with equality of outcomes, that each individual receives a fair share. The pro-social prefers cooperative interactions and will behave as such until she interacts with a consistently noncooperative partner. The second type, the "altruist," acts to promote the welfare of the other above and beyond their own welfare, even in the face of defection. "Pro-selfs" are of two types and are less inclined toward cooperation. In general, pro-selfs are oriented toward maximizing their own outcomes. One type of pro-self disregards others' outcomes. They tend to maximize their own outcomes by not cooperating. The second type of pro-self actively works to minimize others' outcomes. These strictly homo economicus (i.e., ~100% noncooperative) individuals score higher on the Levenson Self-Report Psychopathy Scale [11]. Pro-selfs are one source of noncooperation within dyadic interactions, but they only make up about a third of some sample populations and are not always noncooperative [10, 23, 106–108].

Sometimes, people are noncooperative because they do not like feeling exploited. If individuals feel they have somehow been taken advantage of, they may retaliate or not cooperate to avoid the negative consequences of being used [98, 104, 109–111].

If intentional noncooperation occurs during a reciprocal dyadic interaction (no matter the cause), and TFT is the dominant strategy, noncooperation will increase overall, and it will be difficult or impossible for the two parties to regain cooperation. However, most individuals are not predisposed to begin interactions noncooperatively, or to act noncooperatively in general [6, 11, 29, 106].

In recent years, researchers have realized that noncooperation is not always intentional. Errors, misperceptions, and uncertainty can lead to unintentional noncooperation or perceived defection, which also decrease cooperation [92]. These unintentional causes of noncooperation significantly affect the way that social dilemmas operate.

8.2. Unintentional Noncooperation. Axelrod's tournaments contain some situational elements that are not present in most real-life interactions. In daily life, we rarely have complete information, equivalent payoffs, perfectly executed actions, and clear, unambiguous behaviors. Most real-world dyadic interactions involve uncertain motives, implementation, information, or unequal outcomes. In addition, most dyadic interactions may be occasionally susceptible to what Bendor and colleagues [86] call *exogenous shocks*. These shocks are environmental features that increase uncertainty within the environment, and about the other's actions; they in turn increase the likelihood of attributions that do not foster cooperative behavior. For example, I may show up late to an appointment because of traffic. Depending on my relationship with the person, they may believe me and chalk it up to random chance or, if it were an important business meeting, consider me to be unreliable and not suitable to do business. Situations like this are called "noise" in game theory literature.

Simulations by Bendor and colleagues demonstrate that when any form of uncertainty or "noise" exists during a PDG, TFT fails to maximize outcomes as it did in Axelrod's tournaments. Bendor [85, 112] demonstrated through modeling that TFT was suboptimal in stabilizing high levels of cooperation in a noisy environment. Bendor and colleagues [86], following Axelrod's PDG tournament model but with the addition of noise (defined as random error),

found TFT failed to win the tournament. TFT placed 8th out of 13 strategies determined by average points earned per period. Kollock [87] similarly modeled his simulation on Axelrod's and using Bendor's definition [85, 112] of noise, designed a simulation in which the dichotomous choice PDG was expanded, so that actors could choose from degrees of cooperation (0–10 points per trial) rather than just Cooperate or Defect. Kollock [87] then varied the amount of noise (±1–5) and when it occurred (0–100% of trials). He also found that of the 6 strategies in the tournament, TFT performed poorly against more generous strategies, particularly as the occurrence of noise increased.

The simulation work of Bendor and colleagues [85, 86, 112] and Kollock [87] is supported by behavioral research on PDG variants. Van Lange and colleagues [92] used an extended choice variant of the PDG (give-some game) similar to the task in Kollock's simulation. They found cooperation decreased when negative noise (unintentional noncooperation) was present if participants interacted with a TFT partner. Klapwijk and Van Lange [89] used an innovative parcel delivery task analogous to a sequential PDG. The parcel task involves taking turns delivering packages in a virtual cityscape with closed streets and other disruptions, which signified noise. Longer delivery time increased the sender's cost (€1.40 per second) but also increased the deliverer's profit (€0.60). The study demonstrated that when interacting with a TFT partner, negative noise (both low (25%) or high (50%)) significantly lowered cooperation compared to when noise was absent.

Bendor [85] asserts the classic PDG lacks ecological validity since it contains perfect information and monitoring. *Incomplete information* implies that an individual lacks full information about their interaction partner's previous choices and outcomes. It is not a direct source of unintentional noncooperation, but its existence allows for misinterpretation of data, which reduces cooperation. In Axelrod's tournaments and behavioral studies of the PDG, the availability of perfect information means (mis)attributions of behavior are unlikely [86]. Based on the results of his simulation study, Bendor [85] concluded that imperfect monitoring creates a harsher environment for TFT to maximize outcomes over several trials. Bendor [85] contends that when less than perfect monitoring is possible, this can lead to difficulties in maintaining or fostering cooperation, in part due to less benign attributions. Furthermore, all of the behavioral studies found that benign impressions and attributions of the TFT partner diminished when noise was present [89, 91, 92, 113]. The other person is more likely to be seen as less cooperative and generous, while being viewed as more greedy and selfish. This decrease in benign impressions is more likely to lead to retaliation [92].

Our understanding of the direct impact of (mis) attribution on social exchange choices is unfortunately limited and deserves further investigation. Even so, the decline in benign impressions when negative noise is present suggests that *misattribution due to incomplete information can cause unintentional noncooperation*. In most real-world settings, perfect information is rare and individuals struggle to understand the behavior of others through attributions. Unfortunately, attributions rarely discern the true motivation or cause of a particular choice and are often less than benign [114–116]. If I have invited a friend over for dinner a couple of times but have yet to be invited over to her house for dinner, I can attribute this lack of an invite to many factors--she is busy, inconsiderate, or stingy. The attribution I make regarding her behavior will likely shape my behavior toward her during our next interaction. Unfortunately, these factors are likely to increase the occurrence of noncooperation, which in a strictly reciprocal environment can threaten outcomes and continued interaction. The situation becomes more uncertain if two individuals lack perfect information or monitoring about the previous behavior of the other or the other's outcomes.

Another source of uncertainty occurs when choices are truly *continuous*, and not dichotomous. Kollack [87] argued that the classic PDG used in most laboratory studies of dyadic interactions lacked a key element present in most real-life interactions, namely the freedom to choose degrees of cooperation, as opposed to absolute choices of either cooperation or defection. Continuous dilemmas are situations in which individuals have a potentially countless range of choices, which can be hard to reciprocate accurately and can increase uncertainty [117].

The problem of continuous dilemmas in dyads has not been adequately investigated, but studies involving continuous resource dilemmas in groups and an experiment on force reciprocation within dyads provide some hint at the difficulties that may arise for participants in continuous dilemmas. In experimental game theory, Herlocker and colleagues [117] examined consumption behavior in a continuous resource dilemma (not a dyadic interaction). They were interested in whether individuals could accurately consume part of a resource (spatial, temporal, or physical) when the choices were continuous. Their findings indicate that individuals' perceptions of resources are not well calibrated. That is, though they wanted to be reciprocal, they took more than they realized. Whereas overconsumption in the noncontinuous dilemma could be attributed only to self-interest, here individuals overconsumed a continuous resource not only due to self-interest but also perceptual errors. Participants were unable to accurately consume a limited portion of the resource due to a perceptual error that resulted in the noncooperative act of overconsumption. If individuals in a dyadic interaction who are following TFT made similar perceptual errors (either in judging their own or the other's behavior), over time this would lead to an increase in noncooperation, a difficult cycle to break.

Neuroscientists interested in how physical conflicts often rapidly escalate examined whether there is a neurological factor in force reciprocation that gives rise to increasing physical force. Shergill and colleagues instructed dyads to reciprocate the same force finger press via a force transducer to one another; in all cases the force escalated quickly [118]. In a second study, the researchers asked individuals to use their left index finger to exert force on their right index finger and then vice versa, and individuals consistently overestimated the amount of force required to reciprocate [118]. Perceptual errors in reciprocity are just one more

realistic source of noise during dyadic interactions, and our understanding of how perceptual error influences choices in continuous dyadic interactions is virtually nonexistent.

9. Beyond Reciprocity

An individual enters into a dyadic interaction with an expectation of positive outcomes and acts in accordance with his/her desired outcomes [26, 54, 59]. During the initial stages of relationship development, individuals use their experiences to forecast future outcomes and interaction norms are established, guiding the choices and behaviors of each individual [59]. In addition, each individual seeks an understanding of what they can expect from the other; strict reciprocity is a useful initial strategy to meet these goals [71]. If outcomes or the mutual rate of cooperation falls below expected, individuals may choose to act noncooperatively and/or exit the interaction during this stage. However, once the pair establishes a pattern of cooperative behavior and invests in the relationship, noncooperation can have different effects. It may lead to uncertainty about whether outcomes will continue to meet expectations and remain stable. Individuals have three choices when faced with a noncooperative act after a pattern of cooperation has been established: respond in kind to retaliate (i.e., TFT), punish (give even less than the other or cause them some difficulty), or forgive (continue at the level of cooperation that was present before the defection).

TFT dictates that the individual responds to noncooperation with noncooperation. Because of this inability to exit a noncooperation spiral, TFT does not serve the purpose of maintaining the dyad. Therefore, a behavioral shift needs to occur in order to restore mutual cooperation. This can come about either through punishment, an attempt to reduce the other's noncooperation through negative consequences, or through forgiveness. Punishment and forgiveness are meant to induce a behavioral shift in the hopes of preventing a negative reciprocity spiral.

10. Cost of Punishment

Punishment does sometimes increase cooperation. It goes beyond strict reciprocity to reduce the other's outcomes by imposing external costs. Punishment signals that the individual cannot be treated badly, which can relieve negative emotions associated with being exploited. The threat of punishment and punishment itself can be used to decrease noncooperation. From a purely homo economicus perspective, if an individual is solely interested in increasing material outcomes and minimizing the negative consequences of being used, punishment can be an effective strategy. However, there are economic, emotional, and social consequences for using punishment during a dyadic interaction that can be detrimental to the continuation of the relationship and outcomes.

10.1. Economic Costs. Axelrod's tournaments [71] indicate that of the nice strategies that did well in the first tournament, those involving the least forgiving performed most poorly. In addition, noncooperation and retaliation had an overwhelmingly negative effect on outcomes, as retaliation breeds more retaliation, creating what Axelrod termed *echo effects*. In the simulations conducted by Bendor and colleagues [86] and Kollock [87], namely punishing (unforgiving) strategies, VIGILANT, TFT-1, and TF2Tmin, were not as successful at maximizing outcomes as TFT or more forgiving strategies.

Rand and colleagues [119], interested in whether punishment reestablishes cooperation after defection and increases outcomes, also conducted a PDG simulation in which three choices were possible: to cooperate, retaliate, or punish. Based on the results of the simulation, they concluded that costly punishment does not promote the evolution of cooperation in a PDG. In a repeated PDG behavioral study, Dreber and colleagues [120] found costly punishment was unable to restore cooperation, and average payoffs were lower when punishment was used compared to TFT style strategies, where defection is responded to with defection and not punishment.

10.2. Social/Emotional Costs. Unfortunately, game theory research on punishment focuses mainly on immediate outcomes or short-term increases in cooperation [121, 122] and overlooks the emotional and long-term interpersonal consequences of using punishment. While these studies on short-term punishment consequences constitute important research, many of the most impactful relationships (social and business) include emotions and require some level of long-term cooperative stability. Therefore, understanding the emotional and long-term social impact of punishment is an important gap in the game theory literature.

Bandura [123] argues that punishment reduces the probability of maintaining an ongoing interaction because it increases the likelihood of a negative psychological reaction to the punisher, i.e., resentment or anger. In a multinational survey study, Strimling and Eriksson [124] compared individuals' perceptions of voluntary punishers versus nonpunishers within a group. Respondents wanted to spend less time with punishers and perceived punishers as angrier and less trustworthy than nonpunishers. Respondents also felt that punishers created worse morale than nonpunishers. Further research by Eriksson and colleagues [125, 126] has demonstrated that individuals do not view active punishment (the reduction of outcomes) to be an appropriate moral action to take in response to an unequal split of an ultimatum game payoff.

Not only do people dislike others who punish, but people avoid punishing because they would dislike themselves. Consistent with the findings of Strimling and Eriksson, Rumble [127] measured the perception of punishers within a dyadic carpooling dilemma that resembles a sequential PDG. The carpooling dilemma describes a situation in which two students take turns driving each other to university in order to save money on parking and fuel. One member of the dyad has not picked up the other (1 day or 4 days or 16 days out of 36 days), and the other member could forgive or punish (refuse to drive for 7 or 14 days). When the other member chose to punish, regardless of level of noncooperation,

participants had a less positive impression of the punisher and were less likely to view the behavior as appropriate and fair. In a second study, when individuals interacting with an uncooperative other (TFT-1,TFT-5, or All Defect) during an extended choice PDG were given a chance to punish without cost (to take away some of the other's outcomes), only 12.5% (8 out of 64 participants) did so, and only 11% (7 out of 64 participants) sent a negative warning or threat. Overwhelmingly, when asked why they chose to send a positive message (31 out of 64) or give the money (18 out of 64), participants expressed the desire to be "a nice guy," "generous," and "kind." This study indicates that individuals are concerned with creating and maintaining a positive self-image and impression, and at least in this situation their investment might not be high enough to warrant damaging either by punishing the other.

In addition to damaging relationships and self-esteem, punishment does not encourage long-term behavior change and therefore requires consistent effort and sacrifice on the part of the punisher. In seeking to understand how to shape animal and human behavior, Skinner [128] examined the influence of punishment as a method of decreasing unwanted behaviors. He ultimately argued that punishment led to long-term avoidance of the aversive stimuli, not to long-term behavioral change. Increases in cooperation as a result of punishment seem to arise as a method for avoiding repeated punishment, but once the threat of punishment is absent, noncooperation returns and continues. For example, people slow down on the highway when there is a policeman nearby but speed up again when consequences are no longer a major concern [129–131]. Similarly, corporal punishment for children can sometimes cause immediate compliance, but is often followed by a host of negative consequences such as increased aggression, increased antisocial behavior, and lower levels of internalized morality [132]. While not a dyadic interaction, research on public goods dilemmas suggests that individuals sometimes seek evasion when punishment is possible [133]. So, if one hopes to encourage cooperative behavior without expending the effort to have a constant threat of punishment, then the evidence suggests that punishment is not an effective strategy.

Klapwijk and Van Lange [89] hint at other social/emotional consequences of punishing strategies. They found that those who use a stingy strategy (i.e., less nice than TFT) were seen as less trustworthy and less moral than those using strict TFT. In addition, messages of anger about PDG choices can lead to reciprocated anger and a less positive impression of the message sender [134].

The costs of punishment are likely different in dyads and larger groups. Yamagishi [48] asserts that the interdependency possible within dyads is not necessarily possible in larger groups. This may lead to different behaviors and motivations to maintain cooperation over time, including differences in how sanctions are used. For example, the communication of anger (which could be seen as a threat of punishment) during PDG is less likely to increase cooperation than messages of disappointment [134]. In groups of three, network reciprocity increases cooperation, but punishment does not [13].

Some behavioral economists argue for the existence of "altruistic punishment" [135]. This is when person A, who is part of a group, punishes another person in the group even though it is costly for person A. Some scientists believe that because this type of punishment is costly, no selfish person would ever punish, and thus making the behavior altruistic [136]. Oddly, while it is currently called "altruistic punishment," it seems to be motivated by negative emotions more than positive ones and therefore might be more accurately termed "costly punishment" [124, 135]. Regardless, evidence shows that in group social dilemmas, the possibility for costly punishment increases the likelihood of cooperation [135].

If costly punishment were altruistically motivated, one would expect a participant who is willing to undergo such a cost to act altruistically in other scenarios. However, in one study, altruistic punishers were no more likely to contribute to the public good than nonpunishers [124]. In that same study, punishers were given four scenarios in which to act. In two scenarios, they could choose to contribute to the public good themselves, with ten trials where they could punish and ten trials where they could not punish. In the other two scenarios they were assigned a partner who contributed to the public good on their behalf, ten trials with and ten without the opportunity to punish. When the participants themselves could punish, they tended to punish when people in the group gave too little. However, when the participants' partners were contributing for them, they also punished their partners when they gave too much. If these participants were altruistically motivated, we would perhaps expect them to punish their partners when they gave too little [124].

Similarly, if costly punishment were altruistic, one would expect that punishment would be reserved for defectors and therefore used to enhance group outcomes. However, participants who were more likely to "altruistically" punish were also more likely to punish their partners when they gave a lot of their money, but not a little of their money [124, 125]. Along these lines, "altruistic" punishment correlates more with trait anger than trait altruism [137]. Socially, the reputation of altruistic punishers can diminish when they punish, depending on the circumstances, while it seems that choosing not to punish does not negatively affect a person's reputation [125, 138]. It seems that many of the disadvantages of punishment are still present even when a punishment is "altruistic."

Although punishment systems may seem like they should encourage cooperation, they in fact sometimes compromise outcomes and cooperation. Individuals who punish must be willing to accept not only the material costs of punishment, but also the social and emotional effects. On the opposite end of the spectrum lies reward. Reward can also promote cooperation and is more likely to do so in long term interactions [139].

11. Forgiveness and Its Advantages

As relationships develop, an individual's motives shift from concern for personal outcomes toward concern for the other and, perhaps even more importantly, toward the

maintenance and continuation of the relationship [26, 54, 59]. Close relationships research has found that at early stages we monitor interaction and our partners' choices to gain insight into their motivations [72]. If our partner demonstrates they are willing to forgo their self-interest, either for our own or the good of the relationship, we are also more likely to do the same, and subsequent interactions become more satisfying for each partner. This marks a shift in behavior away from self-concern toward concern for others, a critical shift in continuing relationships of all varieties [57, 69, 140].

This does not mean that all defection or continued defection is not reciprocated, just that occasional forgiveness and generosity help to maintain the stability of long-term continuing relationships. A behavioral shift to a more generous reciprocal strategy is most likely to maximize outcomes and promote the long-term maintenance of cooperation in a dyadic relationship. This strategy involves forgiving *occasional* noncooperative acts, regardless of whether they are intentional or unintentional, and avoids the Achilles' heel of TFT and the costs of punishment. In Mark Twain's words "*. . .it is better to be generous, and in the end more profitable. . ..*"

In game theory, forgiveness is often defined as "resisting retaliation." This definition is based on the assumption that humans are like homo economicus. If we are all selfish, then our initial impulse to defection would be to retaliate. Forgiveness, then, could only come about through resisting this impulse. However, we have seen that selfishness is not the de-facto motivation or conduct during interactions. As far as we know, there have been no studies that investigate whether or not retaliation is, in fact, the pervasive emotional response to defection.

Retaliation may not be the de facto response to defection. People respond to defection with forgiveness when they feel empathy [102], and when they interact with in-group members [141, 142]. There are certainly times when the impulse to retaliate is absent. Also, punishers are not viewed positively in groups [143], and punishment is more tied to anger than to altruism [137], and cooperation is a common occurrence, even in the face of defection. It seems unlikely, then, that humans evolved in such a way that we have to expend energy in order to maintain such a common thing as forgiveness.

It seems more likely that forgiveness sometimes involves actively resisting retaliation and other times is more natural. However, studies on the motivation of forgiveness seem to be absent in the literature. That is why we prefer a more behavioral definition of forgiveness. "Resisting retaliation" infers this yet unconfirmed internal processes motivating forgiveness. "Returning to previous levels of cooperation" makes fewer assumptions about the forgiver's internal state, while still accurately describing the behavior in question.

Forgiveness is choosing to act generously or cooperatively in the face of defection. Generosity is the act of giving to another person more than they have given you and can occur independent of forgiveness and without expectation of receiving anything in return [92, 102]. Generosity does not have to be very large to affect the interaction; simply continuing to act cooperatively following noncooperation is an act of generosity. For example, if two roommates alternate nights doing the dishes, and one individual skips her night to do dishes, the other roommate can act generously by just doing the dishes the following night rather than skipping a night (retaliation). When an individual forgives the other for noncooperative action and acts generously, she has moved beyond strict reciprocity, TFT. By forgiving and then acting generously, she increases the length and profitability of the relationship more than if she had retaliated or punished. In addition, once a cooperative relationship has been established and concern for the other and the continuation of the relationship take precedence over concern for personal outcomes, moving beyond TFT with forgiveness and generosity are more likely to serve these goals.

Simulations and behavioral research demonstrate that when uncertainty or noise exists in the environment, TFT and punishing strategies do not maximize outcomes [86, 87, 101, 112, 120]. Given that TFT was found to be ineffective in a noisy PDG, the primary goal of the simulation research by Bendor and colleagues [86] was to determine whether there were other strategies better suited for noisy environments. They concluded that more generous strategies, such as Nice and Forgiving (NICE), which would only retaliate if the other's level of cooperation dropped below a maximum cooperation level of 80%, were best suited for a noisy environment. NICE had the highest average per payoff period (17.05), compared to TFT (15.00) and VIGILANT (8.99). In addition, neither TFT nor VIGILANT were able to reduce the echo effects of noise over time and suffered greater reductions in the first half of the interactions due to noise, compared to more generous strategies.

Using his extended choice PDG, Kollock [87] proposed that more relaxed strategies like TFTmax2 or TFT + 1, compared to more restrictive or punitive strategies like TFT, TFTmin2, or TFT-1, would perform better in noisy environments. In fact, as the occurrence of noise increased from 0 to 40%, the difference between TFT2max and TFT + 1 performance increased in comparison to more restrictive or punitive strategies. However, this advantage was lost as noise increased from 80 to 100%, and all strategies performed about the same. The simulation results by Rand and colleagues [119] illustrate that when the cost of cooperation is less than punishment, generous strategies prevail, and when the cost of punishment is higher, cooperative strategies prevail. Regardless, winning strategies did not use costly punishment in an effort to restore cooperation. Results from simulations research support the argument that generous or relaxed reciprocal strategies perform better than punishing strategies in non-noisy environments, and better than TFT and punishment in noisy (more realistic) environments.

Behavioral research by Dreber and colleagues [120] examined the use of costly punishment in a non-noisy PDG and found that punishment performed more poorly than both mutual cooperation (if possible) and turning the other cheek (resisting retaliation and punishment in response to noncooperation). But Van Lange and colleagues' (2004) research on the influence of negative noise in an extended choice PDG provides the most definitive answer to whether a

generosity is preferable to TFT. When participants interacted with a TFT partner and noise was present, cooperation decreased compared to when no noise was present; but when participants interacted with a TFT + 1 partner, cooperation did not decrease when noise was present. Similarly, Klapwijk and Van Lange [89] found that when participants interacted with a generous partner during the parcel delivery task when noise was present, cooperation was higher than when interacting with a TFT partner, or a more stingy partner. This behavioral evidence supports the simulations' evidence that generous or relaxed reciprocal strategies perform better than punishing strategies in non-noisy environments, and better than TFT and punishment in noisy (more realistic) environments.

12. Beyond Reciprocity: Mechanisms That Promote Generosity

Caporael's argument that we are social creatures whose communal behaviors help us adapt to hostile environments stands in direct contrast to the Hobbesian view of every man for himself [4]. Dyadic relationships suffer reduced outcomes and shorter durations if we all follow a purely self-interested Hobbesian or strictly reciprocal strategy. The preceding sections demonstrate that forgiving and acting generously improve outcomes and cooperation more than punishment or retaliation.

Forgiving and acting generously require corresponding motivation. Cialdini [144] argues that when we act for the benefit of another it is because our self-concept has shifted to include the other person. We begin to incorporate the other person, the relationship itself, and the benefits of that relationship within our self-concept. This allows our conception of self-interest to include the other's interests, thus fostering forgiveness and generosity [144]. Batson, on the other hand, suggests that altruistic concern for the other is a common human capability. One's concern for the other individual and the relationship motivates her to forgive and act generously [145]. Whether forgiveness and generosity are selfish or altruistic, they arise based on an adjustment of motivation (except for the small percentage of people who are behaviorally pure altruists).

In their research on one of the most fundamental dyadic relationships, romantic couples, Rusbult and colleagues [58] have used interdependence theory to understand shifts in motivation during the course of a romantic relationship. One aspect of this theorizing addresses how a dyad develops *pro-relationship motivation*, or the motivation to benefit the partner or relationship by resisting retaliation and forgiving. Holmes and Rempel [72] discuss how dyads in the early stages of a relationship face *diagnostic situations,* where their self-interests are in competition. During these situations, commitment and trust are built if each partner is willing to forgo personal interest for the sake of the relationship or the other person.

Rusbult and colleagues [58] contend that forgiveness and resisting retaliation for noncooperation eventually become automatic behaviors. This does not mean retaliation never occurs and is always resisted. In the interest of the relationship continuing and concern for the other, we balance our interests against the interests of the relationship and the other.

Pro-relationship motivation may develop in other types of dyadic interactions over time, but more research is needed to fully explore this dynamic in nonromantic dyads. Research on close relationships and social value orientation have demonstrated that other-oriented or partner-oriented motivations do exist in nonromantic dryads [6, 12, 146], but how they develop and grow is not as well understood in game theory.

12.1. Trust. The development of trust between two individuals helps to maintain cooperation over time. It helps individuals navigate risky situations, such as when outcomes are unequal or when noncooperation occurs. Early in a dyadic interaction, individuals can face *diagnostic situations* in the form of *trust and strain tests* [72, 147]. The first interaction between individuals is a form of *trust test* and occurs when two individuals make choices that forgo immediate self-interest (defection) in favor of mutual cooperation [29]. When individuals forgo the self-interested choice during a *trust test*, it signals to the other a desire to connect and work toward mutually rewarding outcomes [147, 148]. Trust is built over time as individuals make repeated, mutually cooperative choices [147, 148]. TFT helps individuals navigate their *trust test* by ensuring mutual cooperation at the onset of an interaction.

As the interaction develops, dyads may face *strain tests*, in which individual interests are not aligned. In these situations, mutual cooperation has either been violated or cannot resolve the situation. *Strain tests* are conflicts of interest in which trust must be built through more generous behavior. *Strain tests* require that one individual sacrifice personal outcomes in the short term for partner or relationship-based interests to build trust and maintain cooperation [63].

To navigate either trust or strain tests and maintain cooperation, individuals must balance the need to connect with the desire to protect themselves from exploitation [148]. Murray and Holmes' [148] model of mutual responsiveness discusses how romantic couples demonstrate the desire to connect and resist the desire to self-protect, which increases trust and commitment. A similar model could be extended to all dyads, since similar processes occur as any dyadic interaction develops and struggles to maintain cooperation. A central premise of Murray and Holmes' [148] model asserts that trust is central to promoting mutual cooperation and deterring self-protection. Trust develops through interactions that demonstrate the other's willingness to make mutually cooperative choices and resist immediately self-interested choices.

In the Murray and Holmes' model, trust reduces the likelihood of being replaced by another, in that alternative interactions are less attractive. While this is an easy assumption to make when considering exclusive romantic relationships, it also has some applicability in nonromantic relationships. Dyads of all varieties form in part due to some

unique combination of features that may or may not be achievable with an alternative partner. Even if there are possible alternatives, the value of an established mutually cooperative pattern of behavior may reduce attraction to the alternative.

As dyadic interaction develops, individuals form rules (if, then statements) that govern choices during *strain tests* [148]. For example, Lewicki and colleagues [149, 150] argue that the level of trust dictates behavior during these interactions. In deterrence-based trust, individuals are unsure of the other. They rely on behavioral monitoring, retaliation and punishment to ensure cooperation. TFT, or strict reciprocity, is the behavioral marker of deterrence-based trust at the beginning of an interaction. Once individuals move beyond this stage to knowledge-based trust, they are more likely to be tolerant in ambiguous, low risk strain tests and act generously toward momentary noncooperation. They are able to resist retaliation and punishment because their knowledge-based trust allows them to rely on their knowledge of past cooperation to predict that their partner will again be cooperative in the future. However, knowledge-based trust does have its limitations. Individuals may be less likely to cooperate or resist retaliation and punishment in high-risk situations or with continued noncooperation. These types of situations prime self-protective tendencies. If a relationship has higher levels of trust (Identity-based trust), the partner is less likely to self-protect in the face of noncooperation. Because they have internalized the other's wants and needs, they are more likely to forgive or accommodate noncooperation and less likely to retaliate or withdraw. Cooperation in high-risk situations garners goodwill and clearly communicates a desire to connect, but at a potential cost, and is probably not sustainable.

12.2. Empathy. Empathy, another mechanism that promotes generosity in relationships, is adopting a person's perspective and having understanding they are worse off than typical. It has been shown to increase helping behavior. Batson and Ahmad [151] found that in one-shot experimental games, participants in the empathy condition were more likely to be cooperative, even with a noncooperative other, than in the no-empathy condition. Further, Rumble and colleagues [102] found that individuals who were induced to feel empathy for the other person acted more generously and were able to overcome the potential detrimental influence of noise (unintentional noncooperation) in a repeated, extended choice PDG, though it was less successful at motivating generosity when the noncooperation was intentional.

One way empathy may motivate generosity is by changing the types of attributions made regarding a noncooperative act [102]. Benign attributions for noncooperative behavior could lead to more forgiveness and less retaliation, ultimately avoiding negative reciprocity spirals [85, 86, 92]. The influence of attribution type on cooperative behavior is not very well understood, other than as a secondary variable or through indirect evidence and is clearly an area that requires more research.

12.3. Group Membership. Relational-self orientation, or the degree to which an individual's self-identity is defined by his relationships with others, may also influence whether or not an individual is likely to resist retaliation. Chen and colleagues [152] found that reciprocity in situations that was negatively inequitable to the individual depended on the relationship type and relational-self orientation. Individuals who interacted with close others reacted less negatively to inequitable outcomes, and individuals high on relational-self orientation also responded less negatively.

Group membership can also influence whether we respond cooperatively or retaliate to noncooperation or defection [153]. If two in-group members are interacting and one of them defects (acts noncooperatively), the other is less likely to retaliate than if two members of different groups are interacting [153]. When we interact with in-group members, we are likely to assume continuing interactions, and there is a motivation shift away from self-concerns toward group concerns. On the other hand, we are more likely to expect interactions with out-group members to be short-lived and are less likely to shift away from our own concerns. What is interesting about the above study is that although individuals had negative feelings toward the in-group noncooperator, they still cooperated. The finding indicates that motives shift during these interactions, and individuals are willing to incur costs with group members.

In addition, Yamagishi and colleagues have argued that individuals intuitively assume that a system of generalized exchange, in which one's cooperation toward a member of the system is reciprocated by another member, is established in their group [154, 155]. As such, individuals expect that cooperation with other in-group members would be beneficial in the long run and are, thus motivated to do so. Moreover, since other in-group members share the same expectation about within-group cooperation, individuals expect other group members to display cooperation rather than defection. Previous empirical studies have robustly demonstrated that the assumption of the within-group generalized exchange is a strong driver of cooperation [154, 156, 157]. This further suggests that group membership may function as a buffer against social noise; given that in-group cooperation is beneficial, and people expect other in-group members to act cooperatively by default [154], it would be likely that individuals see selfish behavior of an in-group member as unintentional rather than intentional one. Therefore, shared group membership would encourage forgiveness.

12.4. Emotions and Heuristics. Numerous researchers have argued that the role of social or moral emotions in encouraging cooperative behavior has often been overlooked and deserves more research [158–160]. De Francisco [158] contends that social emotions must be considered when examining cooperative behavior, since emotions are involved in moral judgements and are used to make judgements about fairness and reciprocity. DeSteno and colleagues [159] call for a richer understanding of the role of emotions in exchanges. This would not involve

understanding not only how we put off the gratification of fulfilling our immediate self-interest, but also what we gain from the strategic long-term self-interest of cooperative relationships.

Social and/or moral emotions can promote cooperation and generosity beyond a consideration of self-interest. Pro-social emotions, such as empathy, love, or gratitude, can foster cooperation. We may cooperate or act generously with someone for whom we feel empathy even when we lose tangible outcomes. We do this not as part of a strategy to gain later but just because we feel a particular pro-social emotion. DeSteno and colleagues [159] argue that gratitude (and possibly other pro-social emotions) stabilizes exchanges and promotes long-term interactions by fostering cooperation and generosity. Algoe and Haidt [161] suggest that different pro-social emotions have different, but similarly pro-social, effects on behavior. For example, although both are responses to moral excellence, elevation encourages pro-sociality and affiliative behavior, while gratitude emphasizes improving relationships with benefactors. Other social emotions, such as shame, guilt, or anger, may also influence cooperation and must also be understood.

Intuitive cognitive processes may support cooperation in daily life. Heuristics are cognitive short-cuts we use to help us make decisions in a complex world and are beneficial to maneuvering through our daily lives [162]. The social heuristic hypothesis (SHH) contends that under time pressure, cooperation is the default in social interactions for naïve participants (i.e., participants lacking previous exposure to experimental games), not self-interested behaviors [162]. Rand and colleagues [162] maintain that in everyday life we are more likely to react to situations using intuitive processes based on prior experience, as opposed to reflexive, thoughtful processes, and our intuitive processes favor cooperation. The contention of Caporael and colleagues [4, 35] that we evolved in a group-based society where group interest held precedence over pure self-interest supports Rand and colleagues' [162] assertion that we intuitively tend to cooperate rather than act selfishly.

12.5. Culture. Finally, it is worthwhile noting that cultural and ecological factors may navigate responses to defection in dyads. Eriksson et al. [138], for instance, have found that individuals in different cultures have differing perceptions of the appropriateness of different types of punishments. This suggests that there would be cultural variations in how people think they should respond to defections, retaliation, punishment, or forgiveness. Moreover, one of the ecological factors which might be particularly relevant to the issue is relational mobility [163]. Countries and societies of high relational mobility offer individuals an abundance of opportunities to find new people and form new relationships. Contrastingly, individuals in countries low in relational mobility, given a dearth of social opportunities, individuals tend to form closed and committed relationships with others. Thus, individuals often remain in the same relationship. Cultures high and low in relational mobility, thus, substantially vary in the degree to which individuals find it easy to exit an undesirable relationship and start seeking a new one. This suggests that forgiveness, which allows dyads to keep the relationship after defection, may be a likely option for those in societies low in relational mobility. In contrast, those in societies high in relational mobility may be more likely to opt for retaliation and punishment than those in societies low in relational mobility, which leads them to leave the relationship. Therefore, it would be of vital importance to consider the potential cultural variations in order to understand how individuals respond to defection.

13. Toward an Understanding of Dynamic Interactions

Many questions remain concerning the psychological processes that influence the dynamic interactions that dominate our daily lives. Dyadic interactions deserve a more nuanced examination than those of the reductionistic, homo economicus view, and the behavioral economists' outcome-based investigations. Self-interest and material outcomes are just two aspects of any interdependent interaction, but there are many other emotions, cognitive processes, and motivations that shape dyadic relationships. These diverse aspects of our interactions function differently through time. Therefore, timing is another aspect of relationships that needs to be incorporated into our study of their dynamics.

A visit to my local butcher is a dyadic exchange, money for steak. However, that does not describe the full and dynamic nature of our interaction. Yes, I pay money and receive the steak, but I also have a conversation with the butcher that connects me to the exchange and is personally satisfying beyond my purchase. In addition, the butcher can gain through the conversation both personally and materially. She may feel that conversations with customers build social connections that add to her satisfaction with daily life. Also, good conversation, a high-quality product, and generous portions may cause me to promote her business (which remains unknown to the butcher). Both the butcher and I gain from the interaction in both tangible and intangible ways, even within an uncertain environment. Do I expect to receive exactly the same service and conversation each time I shop at the butcher? Certainly not. At times the butcher will be too busy to talk. Do I stop going to this butcher because of this? Certainly not. Over the course of the relationship, I look at both the tangible and intangible outcomes and assess whether or not I should continue to patronize this butcher. An interaction is just that: interactive and dynamic. As we interact with each other, the butcher also undergoes similar psychological processes and reacts to me.

Up to this point in our paper (and in the game theory literature), we have focused on single strategies to model human interactions. This is a fundamental limitation to seeking a more accurate understanding of relationship dynamics. Dyadic interactions in the real world rarely rely on a single strategy to guide behavior. Instead, dyads move through interaction styles, seeking to balance competing motivations and demands by using dynamic strategies that evolve over time. Here is a brief look at what research

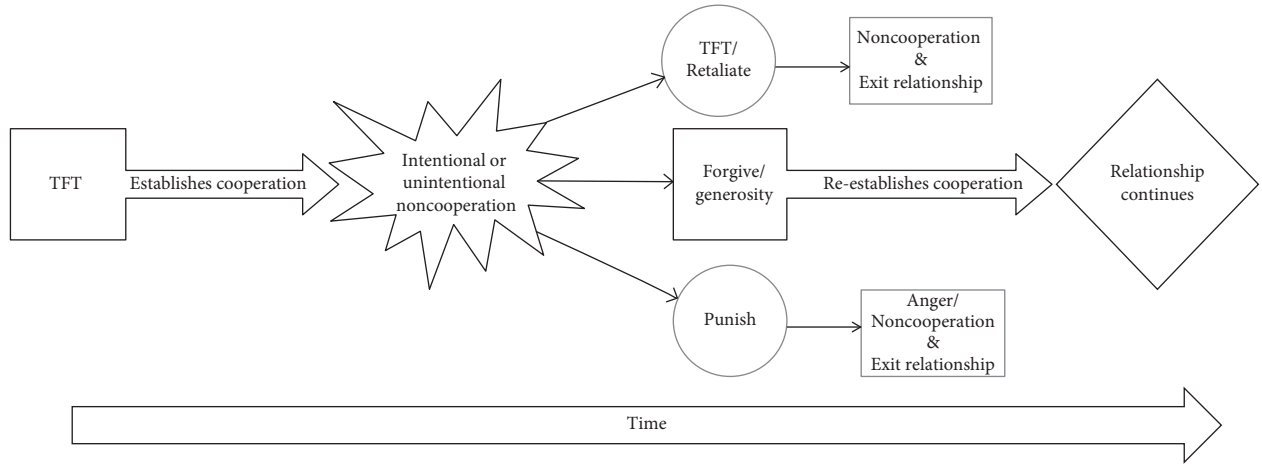

Figure 1: Model of Dyad interactions.

suggests about how relationships shift over time (see Figure 1).

13.1. Beginning. Strict reciprocity, TFT, is particularly useful at the onset of interactions because it simultaneously encourages cooperation and protects against exploitation. TFT helps dyads build mutual cooperation at the beginning of the interaction. Mutual cooperation helps dyads build trust, successfully negotiate *trust tests*, and begins to shift motivations from self and self-protection to the other and the interaction. But if noncooperation occurs during this stage, it is more likely to be reciprocated to avoid exploitation, since trust is nascent. A noncooperative spiral will likely occur, and both parties may choose to exit the interaction.

13.2. Middle. Once a pattern of mutual cooperation and knowledge-based trust has developed, individuals will be more motivated to continue the interaction, and alternatives will appear less attractive. But as the interaction develops, the risks of noncooperation and conflict of interest increase. To reestablish/maintain cooperation, dyad members can choose from several strategies. They can continue to use TFT, though evidence suggests that TFT is unlikely to encourage cooperation in the face of noncooperation. They can punish noncooperation, but they must consider the material, emotional and psychological consequences. They could also act noncooperatively (possibly motivated by self-protection) when facing conflicts of interest, which would also undermine continued cooperation. Finally, individuals can rely on partner/relationship focused emotions to help shift the dyad to a more forgiving, generous strategy that accommodates occasional noncooperation and increases the likelihood of continued cooperation.

A more generous, dynamic, reciprocal strategy allows for occasional noncooperation, and involves reciprocal sacrifice and continued cooperation. Dyads who were able to establish a new mutually cooperative interaction using TFT will be able to maintain that valued interaction through this dynamic, generous strategy. Generous reciprocal strategies maintain cooperation in the face of noncooperation and conflicts of interest. They also prevent exploitation by only resisting retaliation for occasional noncooperation, thus not being a "sucker."

13.3. End. If an interaction continues long term, the individuals may face periods of noncooperation due to shifting circumstances, or there are signals the relationship is ending. How long an individual is willing to tolerate extended noncooperation depends on her commitment to the relationship, the accessibility and quality of alternative interactions, and attributions for the noncooperation. If commitment is weak, the individual will be more motivated to retaliate or exit the interaction. Attractive alternatives may also lead the individual to exit the interaction or retaliate (although this may be resisted to avoid social impression consequences) [76]. If few alternatives exist, a combination of retaliation and punishment may be used, not necessarily to reestablish cooperation, but to assuage negative feelings associated with being duped [98]. Benign attributions are less likely to lead to retaliation and punishment than negative attributions. For example, people may be less likely to retaliate if noncooperation is the result of financial or personal hardship than if the other person seems to be acting selfishly. However, long term ongoing noncooperation will not be tolerated, and retaliation will occur. In addition, if one of the partners becomes consistently noncooperative, the other may assess alternative interactions and leave for a better alternative. Or if few alternatives exist, a combination of retaliation and punishment may be used, not necessarily to reestablish cooperation, but to assuage negative feelings associated with being a sucker.

14. Conclusion

Dyadic relationships are dynamic interactions that dominate our daily lives. In this paper, we provided evidence that human interaction is more than a constant search for maximizing outcomes and satisfying self-interest. Most of our social life is spent interacting with a single other individual, and thus the importance of these interactions must be

reflected in close relationships research. The basic structure, dynamics, purpose, function, and challenges of dyads are inherently different from those of groups. The diffusion of responsibility and blame present in groups does not exist in dyads, where accountability is more transparent. What is unclear is not the source of the noncooperation, but the motivation behind it. In situations like this, attributions are particularly important for future interactions.

Social emotions, heuristics, and motivations all guide behavior in dyadic interactions, and appropriate strategies can maintain stability of cooperation to preserve an ongoing, profitable interaction. Self-interest does not dominate our dyadic interactions, and reciprocity and punishment have limited usefulness. Instead, we are creatures who intuitively understand Mark Twain's assertion that to forgive and "act generously will be more profitable. We will gain gratitude and love."

Conflicts of Interest

The authors have no known conflicts of interest to disclose.

Acknowledgments

The authors would like to thank Marilynn Brewer and Norbert Kerr for their advice during various stages of developing the manuscript. In addition, the authors would like to thank Tim Hopthrow and the School of Psychology at the University of Kent for providing both a stimulating environment and resources to work on the manuscript during Ann Rumble's sabbatical. Finally, they thank Katherine Willcox Ozsari for editing assistance.

References

[1] M. Twain, H. E. Smith, and B. Griffin, *Autobiography of Mark Twain*, University of California Press, Berkeley, CA, USA, 2010.

[2] P. Fleming, *The Death of Homo Economicus*, University of Chicago Press Economics Books, Chicago, IL, USA, 2017.

[3] T. Hobbes, *Leviathan*, 1651.

[4] L. R. Caporael, "The evolution of truly social cognition: the core configurations model," *Personality and Social Psychology Review*, vol. 1, no. 4, pp. 276–298, 1997.

[5] H. Gintis, "Beyond homo economicus: evidence from experimental economics," *Ecological Economics*, vol. 35, no. 3, pp. 311–322, 2000.

[6] J. L. Pletzer, D. Balliet, J. Joireman et al., "Social value orientation, expectations, and cooperation in social dilemmas: a meta–analysis," *European Journal of Personality*, vol. 32, no. 1, pp. 62–83, 2018.

[7] R. Dawkins, *The Selfish Gene*, Oxford University Press, Oxford, UK, 1989.

[8] M. L. Myers, *The Soul of Modern Economic Man: Ideas of Self-Interest, Thomas Hobbes to Adam Smith*, University of Chicago Press, Chicago, IL, USA, 1983.

[9] L. A. Stout, *Cultivating Conscience: How Good Laws Make Good People*, Princeton University Press, Princeton, NJ, USA, 2011.

[10] P. A. M. Van Lange, E. M. N. De Bruin, W. Otten, and J. A. Joireman, "Development of prosocial, individualistic, and competitive orientations: theory and preliminary evidence," *Journal of Personality and Social Psychology*, vol. 73, no. 4, pp. 733–746, 1997.

[11] T. Yamagishi, Y. Li, H. Takagishi, Y. Matsumoto, and T. Kiyonari, "In search of Homo economicus," *Psychological Science*, vol. 25, no. 9, pp. 1699–1711, 2014.

[12] J. Crocker, A. Canevello, and A. A. Brown, "Social motivation: costs and benefits of selfishness and otherishness," *Annual Review of Psychology*, vol. 68, no. 1, pp. 299–325, 2017.

[13] X. Li, M. Jusup, Z. Wang et al., "Punishment diminishes the benefits of network reciprocity in social dilemma experiments," *Proceedings of the National Academy of Sciences*, vol. 115, no. 1, pp. 30–35, 2018.

[14] L. A. Imhof, D. Fudenberg, and M. A. Nowak, "Evolutionary cycles of cooperation and defection," *Proceedings of the National Academy of Sciences*, vol. 102, no. 31, pp. 10797–10800, 2005.

[15] S. S. Komorita and C. D. Parks, "Interpersonal relations: mixed-motive interaction," *Annual Review of Psychology*, vol. 46, no. 1, pp. 183–207, 1995.

[16] A. Sánchez-Amaro, S. Duguid, J. Call, and M. Tomasello, "Chimpanzees and children avoid mutual defection in a social dilemma," *Evolution and Human Behavior*, vol. 40, no. 1, pp. 46–54, 2019.

[17] M. Deutsch, Y. Epstein, D. Canavan, and P. Gumpert, "Strategies of inducing cooperation: an experimental study," *Journal of Conflict Resolution*, vol. 11, no. 3, pp. 345–360, 1967.

[18] L. B. Lave, "Factors affecting cooperation in the prisoner's dilemma," *Behavioral Science*, vol. 10, no. 1, pp. 26–38, 1965.

[19] L. Solomon, "The influence of some types of power relationships and game strategies upon the development of interpersonal trust," *Journal of Abnormal and Social Psychology*, vol. 61, no. 2, pp. 223–230, 1960.

[20] C. D. Batson, "Empathy-induced altruistic motivation," in *Prosocial Motives, Emotions, and Behavior: The Better Angels of Our Nature*, M. Mikulincer, P. R. Shaver, M. Mikulincer, and P. R. Shaver, Eds., American Psychological Association, pp. 15–34, Washington, DC, USA, 2010.

[21] L. R. Caporael, R. M. Dawes, J. M. Orbell, and A. J. C. van de Kragt, "Selfishness examined: cooperation in the absence of egoistic incentives," *Behavioral and Brain Sciences*, vol. 12, no. 4, pp. 683–699, 1989.

[22] J. Henrich, R. Boyd, S. Bowles et al., "'Economic man' in cross-cultural perspective: behavioral experiments in 15 small-scale societies," *Behavioral and Brain Sciences*, vol. 28, no. 6, pp. 795–815, 2005.

[23] P. A. M. Van Lange, "The pursuit of joint outcomes and equality in outcomes: an integrative model of social value orientation," *Journal of Personality and Social Psychology*, vol. 77, no. 2, pp. 337–349, 1999.

[24] M. B. Brewer, "Ambivalent sociality: the human condition," *Behavioral and Brain Sciences*, vol. 12, no. 4, p. 699, 1989.

[25] K. Lewin, *Resolving Social Conflicts: Selected Papers on Group Dynamics*, Harper & Row, New York, NY, USA, 1948.

[26] J. W. Thibaut and H. H. Kelley, *The Social Psychology of Groups*, Transaction Publishers, Piscataway, NJ, USA, 1959.

[27] D. A. Urbina and A. Ruiz-Villaverde, "A critical review of homo economicus from five approaches," *The American Journal of Economics and Sociology*, vol. 78, no. 1, pp. 63–93, 2019.

[28] R. M. Dawes, "Social dilemmas," *Annual Review of Psychology*, vol. 31, no. 1, pp. 169–193, 1980.

[29] D. Dunning, J. E. Anderson, T. Schlösser, D. Ehlebracht, and D. Fetchenhauer, "Trust at zero acquaintance: more a matter of respect than expectation of reward," *Journal of Personality and Social Psychology*, vol. 107, no. 1, pp. 122–141, 2014.

[30] P. A. Van Lange, J. Joireman, C. D. Parks, and E. Van Dijk, "The psychology of social dilemmas: a review," *Organizational Behavior and Human Decision Processes*, vol. 120, no. 2, pp. 125–141, 2013.

[31] G. Stratou, R. Hoegen, G. Lucas, and J. Gratch, "Investigating gender differences in temporal dynamics during an iterated social dilemma: an automatic analysis using networks," in *Proceedings of the 2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*, pp. 531–536, IEEE, San Antonio, TX, USA, October 2017.

[32] R. L. Moreland, "Are dyads really groups?" *Small Group Research*, vol. 41, no. 2, pp. 251–267, 2010.

[33] M. B. Brewer and L. R. Caporael, "An evolutionary perspective on social identity: revisiting groups," in *Evolution and Social Psychology*, M. Schaller, J. A. Simpson, D. T. Kenrick, M. Schaller, J. A. Simpson, and D. T. Kenrick, Eds., pp. 143–161, Psychosocial Press, Madison, CT, USA, 2006.

[34] L. R. Caporael, "Parts and wholes: the evolutionary importance of groups," in *Individual Self, Relational Self, Collective Self*, C. Sedikides, M. B. Brewer, C. Sedikides, and M. B. Brewer, Eds., pp. 241–258, Psychology Press, New York, NY, USA, 2001.

[35] L. R. Caporael, "Evolution, groups, and scaffolded minds," in *Developing Scaffolds in Evolution, Culture, and Cognition*, L. R. Caporael, J. R. Griesemer, W. C. Wimsatt, L. R. Caporael, J. R. Griesemer, and W. C. Wimsatt, Eds., pp. 57–76, The MIT Press, Cambridge, MA, USA, 2014.

[36] M. B. Brewer, "On the social origins of human nature," in *The Message of Social Psychology Perspectives on Mind in Society*, C. McGarty, S. A. Haslam, C. McGarty, and S. A. Haslam, Eds., pp. 54–62, Blackwell Publishing, Malden, MA, USA, 1997.

[37] P. E. Smaldino, "The cultural evolution of emergent group-level traits," *Behavioral and Brain Sciences*, vol. 37, no. 3, pp. 243–254, 2014.

[38] A. Aktipis, L. Cronk, J. Alcock et al., "Understanding cooperation through fitness interdependence," *Nature Human Behaviour*, vol. 2, no. 7, pp. 429–431, 2018.

[39] J. Gowdy and L. Krall, "The economic origins of ultrasociality," *Behavioral and Brain Sciences*, vol. 39, p. e92, 2015.

[40] P. E. Smaldino, "Social identity and cooperation in cultural evolution," *Behavioural Processes*, vol. 161, pp. 108–116, 2019.

[41] J. Grujic, B. Eke, A. Cabrales, J. A. Cuesta, and A. Sanchez, "Three is a crowd in iterated prisoner's dilemmas: experimental evidence on reciprocal behavior," *Scientific Reports*, vol. 2, no. 1, p. 638, 2012.

[42] V. Capraro and H. Barcelo, "Group size effect on cooperation in one-shot social dilemmas II: curvilinear effect," *PLoS One*, vol. 10, no. 7, Article ID e0131419, 2015.

[43] S. Huck, H. T. Normann, and J. Oechssler, "Two are few and four are many: number effects in experimental oligopolies," *Journal of Economic Behavior & Organization*, vol. 53, no. 4, pp. 435–446, 2004.

[44] J. Wu, D. Balliet, L. S. Peperkoorn, A. Romano, and P. A. M. Van Lange, "Cooperation in groups of different sizes: the effects of punishment and reputation-based partner choice," *Frontiers in Psychology*, vol. 10, p. 2956, 2020.

[45] R. Albanese and D. D. Van Fleet, "Rational behavior in groups: the free-riding tendency," *Academy of Management Review*, vol. 10, no. 2, pp. 244–255, 1985.

[46] M. B. Brewer and R. M. Kramer, "Choice behavior in social dilemmas: effects of social identity, group size, and decision framing," *Journal of Personality and Social Psychology*, vol. 50, no. 3, pp. 543–549, 1986.

[47] J. W. Sweeney, "An experimental investigation of the free-rider problem," *Social Science Research*, vol. 2, no. 3, pp. 277–292, 1973.

[48] T. Yamagishi and K. Sato, "Motivational bases of the public goods problem," *Journal of Personality and Social Psychology*, vol. 50, no. 1, pp. 67–73, 1986.

[49] K. Garner and M. Deutsch, "Cooperative behavior in dyads: effects of dissimilar goal orientations and differing expectations about the partner," *Journal of Conflict Resolution*, vol. 18, no. 4, pp. 634–645, 1974.

[50] X. B. Arriaga, *An Interdependence Theory Analysis of Close Relationships*, 2013.

[51] M. B. Eberly, E. C. Holley, M. D. Johnson, and T. R. Mitchell, "It's not me, it's not you, it's us! an empirical examination of relational attributions," *Journal of Applied Psychology*, vol. 102, no. 5, pp. 711–731, 2017.

[52] C. E. Rusbult, C. R. Agnew, and X. B. Arriaga, "The investment model of commitment processes," in *Handbook of Theories of Social Psychology*, P. M. Van Lange, A. W. Kruglanski, E. T. Higgins, P. M. Van Lange, A. W. Kruglanski, and E. T. Higgins, Eds., pp. 218–231, Sage Publications Ltd, Thousand Oaks, CA, USA, 2012.

[53] P. A. Van Lange and D. Balliet, *Interdependence Theory*, 2015.

[54] P. M. Van Lange and C. E. Rusbult, "Interdependence theory," in *Handbook of Theories of Social Psychology*, P. M. Van Lange, A. W. Kruglanski, E. T. Higgins, P. M. Van Lange, A. W. Kruglanski, and E. T. Higgins, Eds., pp. 251–272, Sage Publications Ltd, Thousand Oaks, CA, USA, 2012.

[55] L. V. Machia and B. G. Ogolsky, "The reasons people think about staying and leaving their romantic relationships: a mixed-method analysis," *Personality and Social Psychology Bulletin*, vol. 47, no. 8, Article ID 014616722096690, 2020.

[56] C. E. Rusbult, "A longitudinal test of the investment model: the development (and deterioration) of satisfaction and commitment in heterosexual involvements," *Journal of Personality and Social Psychology*, vol. 45, no. 1, pp. 101–117, 1983.

[57] C. E. Rusbult and C. R. Agnew, "Prosocial motivation and behavior in close relationships," in *Prosocial Motives, Emotions, and Behavior: The Better Angels of Our Nature*, M. Mikulincer, P. R. Shaver, M. Mikulincer, and P. R. Shaver, Eds., American Psychological Association, pp. 327–345, Washington, DC, USA, 2010.

[58] C. E. Rusbult, V. L. Bissonnette, X. B. Arriaga, and C. L. Cox, "Accommodation processes during the early years of marriage," in *The Developmental Course of Marital Dysfunction*, T. N. Bradbury, Ed., Cambridge University Press, pp. 74–113, New York, NY, USA, 1998.

[59] H. H. Kelley, J. G. Holmes, N. L. Kerr, H. T. Reis, C. E. Rusbult, and P. M. Van Lange, *An Atlas of Interpersonal Situations*, Cambridge University Press, New York, NY, USA, 2003.

[60] A. C. Rumble, "The dimension of time in interdependence theory," in *Understanding Behavior in the Context of Time: Theory, Research, and Application*, A. Strathman, J. Joireman,

A. Strathman, and J. Joireman, Eds., pp. 207–224, Lawrence Erlbaum Associates Publishers, Mahwah, NJ, USA, 2005.

[61] M. A. Korsgaard, H. H. Brower, and S. W. Lester, "It isn't always mutual: a critical review of dyadic trust," *Journal of Management*, vol. 41, no. 1, pp. 47–70, 2015.

[62] J. K. Rempel, J. G. Holmes, and M. P. Zanna, "Trust in close relationships," *Journal of Personality and Social Psychology*, vol. 49, no. 1, pp. 95–112, 1985.

[63] S. L. Shallcross and J. A. Simpson, "Trust and responsiveness in strain-test situations: a dyadic perspective," *Journal of Personality and Social Psychology*, vol. 102, no. 5, pp. 1031–1044, 2012.

[64] H. H. Kelley, "The theoretical description of interdependence by means of transition lists," *Journal of Personality and Social Psychology*, vol. 47, no. 5, pp. 956–982, 1984.

[65] H. H. Kelley, "Expanding the analysis of social orientations by reference to the sequential–temporal structure of situations," *European Journal of Social Psychology*, vol. 27, no. 4, pp. 373–404, 1997.

[66] R. S. Cropanzano, K. M. Kacmar, and D. P. Bozeman, "The social setting of work organizations: politics, justice, and support," *Organizational Politics, Justice, and Support: Managing the Social Climate of the Workplace*, Greenwood Publishing Group, vol. 144, p. 118, Oxford, UK, 1995.

[67] D. T. Miller, "Disrespect and the experience of injustice," *Annual Review of Psychology*, vol. 52, no. 1, pp. 527–553, 2001.

[68] J. Persky, "Retrospectives: the ethology of *Homo economicus*," *The Journal of Economic Perspectives*, vol. 9, no. 2, pp. 221–231, 1995.

[69] S. M. Drigotas and C. E. Rusbult, "Should I stay or should I go? a dependence model of breakups," *Journal of Personality and Social Psychology*, vol. 62, no. 1, pp. 62–87, 1992.

[70] D. Balliet, J. M. Tybur, and P. A. M. Van Lange, "Functional interdependence theory: an evolutionary account of social situations," *Personality and Social Psychology Review*, vol. 21, no. 4, pp. 361–388, 2017.

[71] R. Axelrod, *The Evolution of Cooperation*, Basic Books, New York, NY, USA, 1984.

[72] J. G. Holmes and J. K. Rempel, "Trust in close relationships," in *Close Relationships*, C. Hendrick, Ed., pp. 187–220, Sage Publications, Inc, Thousand Oaks, CA, USA, 1989.

[73] M. L. Visserman, F. Righetti, E. A. Impett, D. Keltner, and P. A. M. Van Lange, "It's the motive that counts: perceived sacrifice motives and gratitude in romantic relationships," *Emotion*, vol. 18, no. 5, pp. 625–637, 2018.

[74] S. Columbus, F. Righetti, and D. Balliet, "Situations in close relationships," in *Interdependence, Interaction, and Close Relationships*, L. V. Machia, C. R. Agnew, and X. B. Arriaga, Eds., pp. 11–36, Cambridge University Press, 2020.

[75] S. L. Murray and J. G. Holmes, "Maintaining mutual commitment in the face of risk," *Current Opinion in Psychology*, vol. 1, pp. 57–60, 2015.

[76] C. E. Rusbult, J. Verette, G. A. Whitney, L. F. Slovik, and I. Lipkus, "Accommodation processes in close relationships: theory and preliminary empirical evidence," *Journal of Personality and Social Psychology*, vol. 60, no. 1, pp. 53–78, 1991.

[77] J. Wieselquist, C. E. Rusbult, C. A. Foster, and C. R. Agnew, "Commitment, pro-relationship behavior, and trust in close relationships," *Journal of Personality and Social Psychology*, vol. 77, no. 5, pp. 942–966, 1999.

[78] L. C. Becker, *Reciprocity*, Routledge & Kegan Paul, London, UK, 1986.

[79] M. R. Firmansyah, F. Faturochman, and W. M. Minza, "Do closeness, support, and reciprocity influence trust in friendship?" *Jurnal Psikologi Sosial*, vol. 19, no. 1, pp. 59–68, 2021.

[80] J. M. Burger, J. Sanchez, J. E. Imberi, and L. R. Grande, "The norm of reciprocity as an internalized social norm: returning favors even when no one finds out," *Social Influence*, vol. 4, no. 1, pp. 11–17, 2009.

[81] R. B. Cialdini, B. L. Green, and A. J. Rusch, "When tactical pronouncements of change become real change: the case of reciprocal persuasion," *Journal of Personality and Social Psychology*, vol. 63, no. 1, pp. 30–40, 1992.

[82] S. Sprecher, S. Treger, J. D. Wondra, N. Hilaire, and K. Wallpe, "Taking turns: reciprocal self-disclosure promotes liking in initial interactions," *Journal of Experimental Social Psychology*, vol. 49, no. 5, pp. 860–866, 2013.

[83] S. F. Brosnan, F. M. de Waal, and D. Proctor, "Reciprocity in primates," in *The Interdisciplinary Science of Consumption*, S. D. Preston, M. L. Kringelbach, B. Knutson, S. D. Preston, M. L. Kringelbach, and B. Knutson, Eds., pp. 3–31, MIT Press, Cambridge, MA, USA, 2014.

[84] A. Falk and U. Fischbacher, "A theory of reciprocity," *Games and Economic Behavior*, vol. 54, no. 2, pp. 293–315, 2006.

[85] J. Bendor, "Uncertainty and the evolution of cooperation," *Journal of Conflict Resolution*, vol. 37, no. 4, pp. 709–734, 1993.

[86] J. Bendor, R. M. Kramer, and S. Stout, "When in doubt: cooperation in a noisy prisoner's dilemma," *Journal of Conflict Resolution*, vol. 35, no. 4, pp. 691–719, 1991.

[87] P. Kollock, "'An eye for an eye leaves everyone blind': cooperation and accounting systems," *American Sociological Review*, vol. 58, no. 6, pp. 768–786, 1993.

[88] W. Tung Au and S. S. Komorita, "Effects of initial choices in the prisoner's dilemma," *Journal of Behavioral Decision Making*, vol. 15, no. 4, pp. 343–359, 2002.

[89] A. Klapwijk and P. A. M. Van Lange, "Promoting cooperation and trust in 'noisy' situations: the power of generosity," *Journal of Personality and Social Psychology*, vol. 96, no. 1, pp. 83–103, 2009.

[90] S. S. Komorita, J. A. Hilty, and C. D. Parks, "Reciprocity and cooperation in social dilemmas," *Journal of Conflict Resolution*, vol. 35, no. 3, pp. 494–518, 1991.

[91] M. J. A. Tazelaar, P. A. M. Van Lange, and J. W. Ouwerkerk, "How to cope with 'Noise' in social dilemmas: the benefits of communication," *Journal of Personality and Social Psychology*, vol. 87, no. 6, pp. 845–859, 2004.

[92] P. A. M. Van Lange, J. W. Ouwerkerk, and M. J. A. Tazelaar, "How to overcome the detrimental effects of noise in social interaction: the benefits of generosity," *Journal of Personality and Social Psychology*, vol. 82, no. 5, pp. 768–780, 2002.

[93] M. Patchen, "Strategies for eliciting cooperation from an adversary," *Journal of Conflict Resolution*, vol. 31, no. 1, pp. 164–185, 1987.

[94] D. Kraines and V. Kraines, "Pavlov and the prisoner's dilemma," *Theory and Decision*, vol. 26, no. 1, pp. 47–79, 1989.

[95] R. B. Lount Jr., C. B. Zhong, N. Sivanathan, and J. K. Murnighan, "Getting off on the wrong foot: the timing of a breach and the restoration of trust," *Personality and Social Psychology Bulletin*, vol. 34, no. 12, pp. 1601–1612, 2008.

[96] R. Montal-Rosenberg and S. Moran, "Envy and help giving," *Journal of Personality and Social Psychology*, 2020.

[97] C. D. Parks, A. C. Rumble, and D. C. Posey, "The effects of envy on reciprocation in a social dilemma," *Personality and Social Psychology Bulletin*, vol. 28, no. 4, pp. 509–520, 2002.

[98] K. D. Vohs, R. F. Baumeister, and J. Chin, "Feeling duped: emotional, motivational, and cognitive aspects of being exploited by others," *Review of General Psychology*, vol. 11, no. 2, pp. 127–141, 2007.

[99] D. A. Effron and D. T. Miller, "Reducing exposure to trust-related risks to avoid self-blame," *Personality and Social Psychology Bulletin*, vol. 37, no. 2, pp. 181–192, 2011.

[100] L. Cosmides and J. Tooby, "Evolutionary psychology and the generation of culture, part II," *Ethology and Sociobiology*, vol. 10, no. 1-3, pp. 51–97, 1989.

[101] K. Sigmund, *The Calculus of Selfishness*, Princeton University Press, Princeton, NJ, USA, 2016.

[102] A. C. Rumble, P. A. M. Van Lange, and C. D. Parks, "The benefits of empathy: when empathy may sustain cooperation in social dilemmas," *European Journal of Social Psychology*, vol. 40, no. 5, pp. 856–866, 2010.

[103] J. H. Vuolevi and P. A. Van Lange, "Boundaries of reciprocity: incompleteness of information undermines cooperation," *Acta Psychologica*, vol. 141, no. 1, pp. 67–72, 2012.

[104] E. Fehr and K. M. Schmidt, "A theory of fairness, competition, and cooperation," *Quarterly Journal of Economics*, vol. 114, no. 3, pp. 817–868, 1999.

[105] D. M. Messick and C. G. McClintock, "Motivational bases of choice in experimental games," *Journal of Experimental Social Psychology*, vol. 4, no. 1, pp. 1–25, 1968.

[106] R. O. Murphy, K. A. Ackermann, and M. Handgraaf, "Measuring social value orientation," *SSRN Electronic Journal*, vol. 6, no. 8, pp. 771–781, 2011.

[107] E. van Dijk, D. De Cremer, and M. J. Handgraaf, "Social value orientations and the strategic use of fairness in ultimatum bargaining," *Journal of Experimental Social Psychology*, vol. 40, no. 6, pp. 697–707, 2004.

[108] J. W. van Prooijen, D. De Cremer, I. van Beest, T. Ståhl, M. v. Dijke, and P. A. Van Lange, "The egocentric nature of procedural justice: social value orientation as moderator of reactions to decision-making procedures," *Journal of Experimental Social Psychology*, vol. 44, no. 5, pp. 1303–1315, 2008.

[109] W. E. Haley and B. R. Strickland, "Interpersonal betrayal and cooperation: effects on self-evaluation in depression," *Journal of Personality and Social Psychology*, vol. 50, no. 2, pp. 386–391, 1986.

[110] N. L. Kerr, "Motivation losses in small groups: a social dilemma analysis," *Journal of Personality and Social Psychology*, vol. 45, no. 4, pp. 819–828, 1983.

[111] R. Madrigal, M. Wardley, and C. Armstrong Soule, *Duped, Scammed and Suckered: The Development of the Sugrophobia Scale*, ACR North American Advances, 2014.

[112] J. Bendor, "In good times and bad: reciprocity in an uncertain world," *American Journal of Political Science*, vol. 31, no. 3, pp. 531–558, 1987.

[113] C. P. Reinders Folmer, T. Wildschut, D. De Cremer, and P. A. M. van Lange, "Coping with noise in social dilemmas: group representatives fare worse than individuals because they lack trust in others' benign intentions," *Group Processes & Intergroup Relations*, vol. 22, no. 2, pp. 200–214, 2019.

[114] M. Hewstone, "The 'ultimate attribution error'? a review of the literature on intergroup causal attribution," *European Journal of Social Psychology*, vol. 20, no. 4, pp. 311–335, 1990.

[115] L. D. Ross, "The intuitive psychologist and his shortcomings: distortions in the attribution process," *Advances in Experimental Social Psychology*, vol. 10, pp. 173–220, 1977.

[116] C. Stephan, N. R. Presser, J. C. Kennedy, and E. Aronson, "Attributions to success and failure after cooperative or competitive interaction," *European Journal of Social Psychology*, vol. 8, no. 2, pp. 269–274, 1978.

[117] C. E. Herlocker, S. T. Allison, J. D. Foubert, and J. K. Beggan, "Intended and unintended overconsumption of physical, spatial, and temporal resources," *Journal of Personality and Social Psychology*, vol. 73, no. 5, pp. 992–1004, 1997.

[118] S. S. Shergill, P. M. Bays, C. D. Frith, and D. M. Wolpert, "Two eyes for an eye: the neuroscience of force escalation," *Science*, vol. 301, no. 5630, p. 187, 2003.

[119] D. G. Rand, H. Ohtsuki, and M. A. Nowak, "Direct reciprocity with costly punishment: generous tit-for-tat prevails," *Journal of Theoretical Biology*, vol. 256, no. 1, pp. 45–57, 2009.

[120] A. Dreber, D. G. Rand, D. Fudenberg, and M. A. Nowak, "Winners don't punish," *Nature*, vol. 452, no. 7185, pp. 348–351, 2008.

[121] W. Li, J. Chen, and B. Zhou, "Game theory analysis for graded punishment mechanism restraining free riding in p2p networks," in *Proceedings of the 2011 International Symposium on Computer Science and Society*, pp. 262–266, IEEE, Kota Kinabalu, Malaysia, July 2011.

[122] H. Rauhut and M. Junker, "Punishment deters crime because humans are bounded in their strategic decision-making," *The Journal of Artificial Societies and Social Simulation*, vol. 12, no. 3, p. 1, 2009.

[123] A. Bandura, *Social Learning Theory*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1977.

[124] P. Strimling and K. Eriksson, "Regulating the regulation: norms about punishment," in *Reward and Punishment in Social Dilemmas*, P. M. Van Lange, B. Rockenbach, T. Yamagishi, P. M. Van Lange, B. Rockenbach, and T. Yamagishi, Eds., Oxford University Press, pp. 52–69, Oxford, UK, 2014.

[125] K. Eriksson, P. Andersson, and P. Strimling, "When is it appropriate to reprimand a norm violation?: the roles of anger, behavioral consequences, violation severity, and social distance," *Judgment and Decision Making*, vol. 12, no. 4, pp. 396–407, 2017.

[126] K. Eriksson, P. Strimling, P. A. Andersson, and T. Lindholm, "Costly punishment in the ultimatum game evokes moral concern, in particular when framed as payoff reduction," *Journal of Experimental Social Psychology*, vol. 69, pp. 59–64, 2017.

[127] A. C. Rumble, *Punishing the Punisher: How Do We Perceive Those Who Punish?*, 2015.

[128] B. F. Skinner, *Science and Human Behavior*, Macmillan, New York, NY, USA, 1953.

[129] E. O. Ryeng, "The effect of sanctions and police enforcement on drivers' choice of speed," *Accident Analysis & Prevention*, vol. 45, pp. 446–454, 2012.

[130] L. Walter, J. Broughton, and J. Knowles, "The effects of increased police enforcement along a route in London," *Accident Analysis & Prevention*, vol. 43, no. 3, pp. 1219–1227, 2011.

[131] B. R. E. Wright, A. Caspi, T. E. Moffitt, and R. Paternoster, "Does the perceived risk of punishment deter criminally prone individuals? rational choice, self-control, and crime," *Journal of Research in Crime and Delinquency*, vol. 41, no. 2, pp. 180–213, 2004.

[132] E. T. Gershoff, "Corporal punishment by parents and associated child behaviors and experiences: a meta-analytic and theoretical review," *Psychological Bulletin*, vol. 128, no. 4, p. 539, 2002.

[133] L. B. Mulder, E. van Dijk, D. De Cremer, and H. A. M. Wilke, "When sanctions fail to increase cooperation in social dilemmas: considering the presence of an alternative option to defect," *Personality and Social Psychology Bulletin*, vol. 32, no. 10, pp. 1312–1324, 2006.

[134] M. J. Wubben, D. D. Cremer, and E. v. Dijk, "How emotion communication guides reciprocity: establishing cooperation through disappointment and anger," *Journal of Experimental Social Psychology*, vol. 45, no. 4, pp. 987–990, 2009.

[135] E. Fehr and S. Gächter, "Altruistic punishment in humans," *Nature*, vol. 415, no. 6868, pp. 137–140, 2002.

[136] E. Fehr and U. Fischbacher, "The nature of human altruism," *Nature*, vol. 425, no. 6960, pp. 785–791, 2003.

[137] J. Rodrigues, N. Nagowski, P. Mussel, and J. Hewig, "Altruistic punishment is connected to trait anger, not trait altruism, if compensation is available," *Heliyon*, vol. 4, no. 11, Article ID e00962, 2018.

[138] K. Eriksson, P. Strimling, M. Gelfand et al., "Perceptions of the appropriate response to norm violation in 57 societies," *Nature Communications*, vol. 12, no. 1, p. 1481, 2021.

[139] D. Balliet, L. B. Mulder, and P. A. M. Van Lange, "Reward, punishment, and cooperation: a meta-analysis," *Psychological Bulletin*, vol. 137, no. 4, pp. 594–615, 2011.

[140] B. Fehr, C. Harasymchuk, and S. Sprecher, "Compassionate love in romantic relationships: a review and some new findings," *Journal of Social and Personal Relationships*, vol. 31, no. 5, pp. 575–600, 2014.

[141] A. Vaish and J. Oostenbroek, "Preferential forgiveness: the impact of group membership and remorse on preschoolers' forgiveness," *Journal of Experimental Psychology: General*, 2021.

[142] H. Zourrig, J. C. Chebat, and R. Toffoli, "In-group love and out-group hate?" a cross cultural study on customers' revenge, avoidance and forgiveness behaviors," *Journal of Business Research*, vol. 68, no. 3, pp. 487–499, 2015.

[143] H. Ozono and M. Watabe, "Reputational benefit of punishment: comparison among the punisher, rewarder, and non-sanctioner," *Letters on Evolutionary Behavioral Science*, vol. 3, no. 2, pp. 21–24, 2012.

[144] R. B. Cialdini, S. L. Brown, B. P. Lewis, C. Luce, and S. L. Neuberg, "Reinterpreting the empathy–altruism relationship: when one into one equals oneness," *Journal of Personality and Social Psychology*, vol. 73, no. 3, pp. 481–494, 1997.

[145] C. D. Batson, N. Ahmad, D. A. Lishner, J. Tsang, C. R. Snyder, and S. J. Lopez, "Empathy and altruism," *The Oxford Handbook of Hypo-Egoic Phenomena*, pp. 161–174, 2002.

[146] M. Perc, J. J. Jordan, D. G. Rand, Z. Wang, S. Boccaletti, and A. Szolnoki, "Statistical physics of human cooperation," *Physics Reports*, vol. 687, pp. 1–51, 2017.

[147] J. A. Simpson, "Psychological foundations of trust," *Current Directions in Psychological Science*, vol. 16, no. 5, pp. 264–268, 2007.

[148] S. L. Murray and J. G. Holmes, "The architecture of interdependent minds: a motivation-management theory of mutual responsiveness," *Psychological Review*, vol. 116, no. 4, pp. 908–928, 2009.

[149] R. J. Lewicki and B. B. Bunker, "Developing and maintaining trust in work relationships," in *Trust in Organizations: Frontiers of Theory and Research*, R. M. Kramer and T. R. Tyler, Eds., pp. 114–139, SAGE Publications, Thousand Oaks, CA, USA, 1996.

[150] R. J. Lewicki, E. C. Tomlinson, and N. Gillespie, "Models of interpersonal trust development: theoretical approaches, empirical evidence, and future directions," *Journal of Management*, vol. 32, no. 6, pp. 991–1022, 2006.

[151] C. D. Batson and N. Ahmad, "Empathy-induced altruism in a prisoner's dilemma II: what if the target of empathy has defected?" *European Journal of Social Psychology*, vol. 31, no. 1, pp. 25–36, 2001.

[152] Y. R. Chen, X. P. Chen, and R. Portnoy, "To whom do positive norm and negative norm of reciprocity apply? effects of inequitable offer, relationship, and relational-self orientation," *Journal of Experimental Social Psychology*, vol. 45, no. 1, pp. 24–34, 2009.

[153] P. Arora, J. Logg, and R. Larrick, "Acting for the greater good: identification with group determines choices in sequential contribution dilemmas," *Journal of Behavioral Decision Making*, vol. 29, no. 5, pp. 499–510, 2016.

[154] T. Yamagishi, M. Kikuchi, and M. Kosugi, "Trust, gullibility, and social intelligence," *Asian Journal of Social Psychology*, vol. 2, no. 1, pp. 145–161, 1999.

[155] T. Yamagishi and N. Mifune, "Parochial altruism: does it explain modern human group psychology?" *Current Opinion in Psychology*, vol. 7, pp. 39–43, 2016.

[156] T. Yamagishi and T. Kiyonari, "The group as the container of generalized reciprocity," *Social Psychology Quarterly*, vol. 63, no. 2, pp. 116–132, 2000.

[157] T. Yamagishi and N. Mifune, "Does shared group membership promote altruism? fear, greed, and reputation," *Rationality and Society*, vol. 20, no. 1, pp. 5–30, 2008.

[158] A. De Francisco, "Feeling cooperation, being moral," *Rationality and Society*, vol. 26, no. 3, pp. 355–372, 2014.

[159] D. DeSteno, M. Y. Bartlett, J. Baumann, L. A. Williams, and L. Dickens, "Gratitude as moral sentiment: emotion-guided cooperation in economic exchange," *Emotion*, vol. 10, no. 2, pp. 289–293, 2010.

[160] J. S. Lerner, Y. Li, P. Valdesolo, and K. S. Kassam, "Emotion and decision making," *Annual Review of Psychology*, vol. 66, no. 1, pp. 799–823, 2015.

[161] S. B. Algoe and J. Haidt, "Witnessing excellence in action: the 'other-praising' emotions of elevation, gratitude, and admiration," *The Journal of Positive Psychology*, vol. 4, no. 2, pp. 105–127, 2009.

[162] D. G. Rand, A. Peysakhovich, G. T. Kraft-Todd et al., "Social heuristics shape intuitive cooperation," *Nature Communications*, vol. 5, no. 1, pp. 3677–3712, 2014.

[163] M. Yuki and J. Schug, "Relational mobility: a socioecological approach to personal relationships," in *Relationship Science: Integrating Evolutionary, Neuroscience, and Sociocultural Approaches*, O. Gillath, G. Adams, and A. Kunkel, Eds., American Psychological Association, pp. 137–151, Washington, DC, USA, 2012.