Journal of Experimental Psychology: General
2009, Vol. 138, No. 3, 341–350


0096-3445/09/$12.00 DOI: 10.1037/a0016073

# Friend or Foe: Subjective Expected Relative Similarity as a Determinant of Cooperation

Ilan Fischer
University of Haifa

Subjective expected relative similarity (SERS) is a descriptive theory that explains cooperation levels in single-step prisoner's dilemma (PD) games. SERS predicts that individuals cooperate whenever their *subjectively perceived similarity* with their opponent exceeds a situational index, namely the game's *similarity threshold*. A thought experiment and 2 experimental studies illustrate and explore SERS's characteristics, showing that the theory predicts cooperation and competition in single-step PD games under 3 informational structures: (a) clear and transparent similarity cues, (b) experienced similarity, and (c) semantic similarity. The study's findings suggest that perceived similarity and its application in SERS play an important role in the evolution of cooperation underlying both kin and group selection mechanisms.

*Keywords:* conflict, cooperation, prisoner's dilemma, kin selection, group selection

Cooperative behavior is essential for the success of both individuals and groups. From hunter–gatherer societies to nation–states, the ability of individuals to trust one another and to cooperate is an indicator of prosperity and wealth. Cooperating societies are capable of complex social behaviors, such as hunting, trading, and building. They are able to create shared resources that increase their survival prospects and their fitness. In some instances cooperation is an obvious choice for all participants. Two children swinging on a seesaw have a natural motivation to cooperate, as do two porters moving an object too heavy for one person to carry alone. In other cases involving indivisible resources, as in winning a match or owning a piece of land that can sustain no more than a single owner, competition is the natural choice. A third category involves a mixture of both cooperation and competition motivations, and is hence regarded as a mixed-motive interaction or game (Rapoport & Guyer, 1966). In this category, individual motivations are governed by the gains derived from exploiting the opponent or from engaging in mutual cooperation, as well as by the losses incurred by unreciprocated cooperation or mutual defection. Any choice involves both risks and opportunities. Cooperating while confronting a competitive opponent may result in being exploited, yet competing while interacting with a cooperative opponent may result in obtaining the attractive gains of exploitation. The expectations of continued interactions with the same opponent noticeably motivate the choice of cooperation (by inducing both the fear of retaliation following the exploitation of a cooperating opponent and the hope for future gratitude following the appreciation of current cooperation). If, however, the prospects of future encounters with the same opponent are low and negative outcomes have crucial consequences, cooperation becomes a nonrational or even self-defeating strategy.

Correspondence concerning this article should be addressed to Ilan Fischer, Department of Psychology, University of Haifa, Mount Carmel, Haifa 31905, Israel. E-mail: ifischer@psy.haifa.ac.il

To explain why people and other organisms do cooperate despite this grim conclusion and to shed light on the conditions under which people are likely to cooperate, researchers and theoreticians have pursued three distinct routes. The first involves attempting to find a convincing rationale in favor of cooperation as a basic argument embedded within the single meeting of two individuals. The second focuses on the study of repeated interactions, addressing issues such as strategic behavior, learning, and sampling. The third route seeks to identify social preferences and predispositions that transcend the selfish motives of the players. While the study of repeated interactions and the study of social preferences continue to generate very fruitful lines of research, the attempt to justify cooperation from the perspective of the interacting individual in single-step games has reached a dead end. The current article pursues this problematic line of logic and shows how it can be revised to account for cooperation in single interactions that involve motivations to benefit both from mutual cooperation and from exploitation of trusting opponents. The proposed solution is suggested as central to the evolution of cooperation underlying both group and kin selection mechanisms.

Note that the theory proposed in the current article is a descriptive one that attempts to explain and predict human behavior. This new theory builds upon the perception of opponent similarity, which is neither included in the game payoff matrix nor considered in game theoretic analysis. Therefore, the proposed theory can be regarded only as an extension of the normative model contingent upon the addition of a behavioral factor. A distinction should also be made between the analysis of (a) single encounters between two strangers who have no intention to continue their interactions and (b) repeated encounters. Although under some constraints the solution proposed in this article for the single interaction scenario may be adapted to the case of continuing interactions, such adaptation is neither minor nor trivial; and because such a solution is related to a substantial literature of strategic interactions it is not addressed in the current article. Finally, the reader, and especially the experienced student of the prisoner's dilemma, is asked to follow the proposed logic patiently. Indeed at first glance the theory may seem to contradict previous lines of thought. Nevertheless, careful reading of the proposed theory and examination of the experimental work should reveal the theory's consistency and validity and eliminate most if not all discrepancies with previous analyses.

## The Prisoner's Dilemma (PD) Game

Since its introduction by Flood and Dresher (1952), the PD game has become a primary model for the study of conflict and cooperation. The game obtained its name from a story describing two separated prisoners, each given an opportunity either to confess to committing a joint robbery (hence defecting or betraying the other prisoner) or to deny committing the crime (hence cooperating with the other prisoner). Although the story itself is fictitious, its structure has far-reaching implications for the study of conflict and cooperation among individuals (Rapoport & Chammah, 1965), within and among groups (Bornstein, 2004; Dawes, 1980; Hardin, 1968; Macy, 1990, 1991a, 1991b; Schuessler, 1990), firms (Dixit & Nalebuff, 1992), nations, and superpowers (Brams, 1975; Fischer & Suleiman, 1997; Schelling, 1963). It has also become the driving mechanism for many studies exploring the dynamics of the evolution of cooperation (Axelrod, 1980a, 1980b, 1981, 1984; Axelrod & Hamilton, 1981).

The dilemma is described by a $2 \times 2$ payoff matrix that allows each player to choose between a cooperative and a competitive (or defective) move. If both players cooperate, each player obtains the reward ($R$) payoff. If both defect, each player obtains the punishment ($P$) payoff. However, if one player defects while the other cooperates, the defector obtains the temptation ($T$) payoff and the cooperator obtains the sucker's ($S$) payoff, where $T > R > P > S$ (and $R \geq \frac{T+S}{2}$, assuring that sharing the payoffs awarded for uncoordinated choices does not exceed the payoffs obtained by mutual cooperation). Observing the payoff structure of the game (see Table 1), each individual player may realize that the competitive choice of defection is a dominant strategy, because it yields a better payoff regardless of the opponent's choice. (If the opponent chooses to cooperate, cooperation will result in the $R$ payoff, whereas defection will yield the larger $T$ payoff; if the opponent chooses to defect, cooperation will result in obtaining the $S$ payoff, whereas defection will yield the larger $P$ payoff.) By choosing to defect, players protect themselves from exploitation and retain the option to exploit a trusting opponent (Lewis, 1979; Shubik, 1970). Because this is the case for both players, mutual defection is the only Nash equilibrium of the game (meaning that given this mutual choice no player has any motivation to unilaterally alter his choice). Paradoxically, when both players choose to defect, they end up obtaining the $P$ payoff, which is less than the $R$ payoff gained by a mutual choice to cooperate, implying that rational players earn less than non-rational players and hence casting doubt on the notion of rationality itself. This grim conclusion led Luce and Raiffa (1957) to suggest changing the structure of social games "whenever it is inherent in the game situation that the players, in pursuing their own ends, will be forced into a socially undesirable position" (p. 97). It also motivated Hardin (1968) to propose the implementation of a law-enforcing central authority. Interestingly, experimental studies have shown that many human players seem to violate rationality by choosing to cooperate (Frank, Gilovich, & Regan, 1993; McCabe, Rassenti, & Smith, 1996; Sally, 1995). The combination of a deficient rational recommendation, the frequently recognized discrepancy between theory and empirical data, and the immense importance of the PD game in understanding cooperation and competition have together motivated researchers to look for additional explanations of cooperative behavior.

Table 1
*The Prisoner's Dilemma Game for Two Players*

| | Player 2 | |
|---|---|---|
| Player 1 | Cooperate | Defect |
| Cooperate | $R$, $R$ | $S$, $T$ |
| Defect | $T$, $S$ | $P$, $P$ |

*Note.* Payoffs for Player 1 (leftmost in each cell) and Player 2 (right) are denoted as reward ($R$), sucker ($S$), temptation ($T$), and punishment ($P$).

Rapoport and Chammah (1965) showed that PD games differ in their capacity to induce cooperation. The researchers analyzed all 30 interval ratios (i.e., [A – B]/[C – D]) that can be formed from the four parameters ($T$, $R$, $P$, and $S$). Their results showed that only two ratios are independent and that the remaining ones can all be derived from these two ratios ($r_1 = [R - P]/[T - S]$; $r_2 = [R - S]/[T - S]$). Although Rapoport and Chammah's studies were unable to completely corroborate the link between the two ratios and cooperation rates, their analyses provide important insights into the impact of payoff intervals on cooperation rates. Other studies have shown that cooperation is dependent on various perceptual variables, among them the labels of the alternatives, the existence of previous communication, or the extent of anonymity (Sally, 1995). The current article integrates perceptual factors with the calculation of cooperation indices into a theory that explains and predicts cooperation in single-step PD games.

## The Discrete Symmetry Argument

The symmetry argument, also known as "the basic argument in favor of cooperation" (Davis, 1985; Hofstadter, 1985; Rapoport, 1960), asserts that rational players should acknowledge the symmetry in their situation and the similarity of their goals. Hence they should rule out the existence of unilateral moves and consider only similar choices, either mutual defection or mutual cooperation. Rapoport (1960) described this argument as follows:

> My partner is like me. Therefore he is likely to act like me. If I conclude that I should confess, he will probably also conclude the same. If I conclude that I should not confess, this is the way he probably thinks. In the first case, we both get −5 [the punishment payoff]; in the second case, we get +5 [the reward payoff]. This indicates that I personally benefit by not confessing. (p. 175)

A somewhat different formulation by Hofstadter (1985) suggests: "If reasoning guides me to say C, then, as I am no different from anyone else as far as rational thinking is concerned, it will guide everyone to say C" (p. 746). This line of reasoning builds upon logic somewhat similar to Laplace's principle of insufficient reason (Laplace, 1783/1953), which suggests assigning equal probabilities to mutually exclusive states of nature with unknown prospects. While Laplace concluded that strictly uncertain events should be assigned equal probabilities because there is no known reason to favor any other probability distribution, the symmetry argument rules out all possibilities apart from identical choices (because there are no known reasons pointing to informational or cognitive differences among the players). To some extent, the symmetry argument also expresses Schelling's (1963) notion of a "tacit meeting of minds" that occurs when two otherwise independent individuals are considering the same information (though the idea of tacit coordination relates to coordination games with multiple Nash equilibria, whereas the PD game is a mixed-motives game with a unique Nash equilibrium). Howard (1988) implemented the symmetry argument by programming a computer simulation that substitutes actual players with simple automata, or agents, that interact with each other in multiple PD games. The simulation enabled agents to read each other's computer code and identify the extent of similarity between their own and their opponent's programming code. Because the computerized environment contained several duplicates, those agents that used the symmetry argument could indeed identify perfect copies of themselves and hence amass the "reward" payoffs derived from mutual cooperation (see also Danielson, 1992).

In spite of the compelling clarity and simplicity of the symmetry argument, its feasibility remains questionable. Hofstadter (1985), though in favor of the argument,[1] noted that

> to many people, this sounds like a belief in voodoo or sympathetic magic, a vision of a universe permeated by tenuous threads of synchronicity, conveying thoughts from mind to mind like pneumatic tubes carrying messages across Paris, and making people resonate to a secret harmony. (p. 746)

Rapoport (1960) also criticized the argument by asking,

> How can my assumption of what my opponent will do influence what he will actually do? By telepathy? This is clearly absurd, because it implies that if I switch my decision, so will he. But if there is no communication between us, how is this possible? (pp. 238–239)

A different line of thought combines the symmetry argument with players' social preferences. It assumes that similarity may emanate from pre-existing social motivations shared by both players. Sen (1974) described other-regarding preferences that motivate both players to attend to the well-being of their opponents. However, although such motivations are highly important for the understanding of social preferences (Fehr & Gaechter, 2002; Gintis, Bowles, Boyd, & Fehr, 2003), they override the selfish motivations induced by the game payoffs. Players who have concerns for the well-being of others may cooperate in spite of PD payoffs and not because of their desire to maximize their personal outcomes.

## The Continuous Symmetry Argument

Though the discrete version of the symmetry argument fails to account for the emergence of cooperation, a different form of the argument has emerged. I will refer to this as the continuous symmetry argument. This form of the symmetry argument builds upon continuous interpretation of the similarity among the players. Lewis (1979) described the role of uncertain similarity among players by introducing the notion of replicas:

> There are replicas and replicas. Some are the same sort of things that I am, others are less so. A flesh-and-blood duplicate made by copying me atom for atom would be one good sort of replica. A working scale model of me, smaller perhaps by a ratio of 1:48, also might serve. So might a pattern of bits in a computer, or beads on an abacus, or marks on paper, or neuron firings in a brain, even though these things are unlike me and replicate me only by way of some complicated isomorphism . . . . The most readily available sort of replica of me is simply another person, . . . . For instance: you, my fellow prisoner. Most likely you are not a very exact replica of me and your choice is not a very reliable predictive process for mine. (p. 238)

This continuous formulation of the symmetry argument is sharpened by Campbell (1985), who explained that the symmetry argument should not be assumed to be identical to complete dependence between the players. In fact, dependence between players can range anywhere from complete independence to complete dependence. This continuous approach is also consistent with Gauthier's (1985) description of translucent players whose disposition to cooperate or to defect may be ascertained with less than certainty by others. All these formulations of the continuous symmetry argument make players' choices conditional upon the extent of similarity rather than on its mere existence. They provide a generalized symmetry argument that includes the discrete form of the argument as one of two specific boundary conditions, described by the extremes of absolute similarity and absolute lack of similarity. But the continuous symmetry argument does not point to specific similarity characteristics that are, more or less, important for the strategic encounter. Moreover, it does not allow one to make actual predictions, account for individual differences, distinguish between situations, or define the minimal extent of similarity that renders cooperation worthwhile. Hence, it cannot be regarded as a complete theory.

A somewhat different perspective regarding continuous similarity was suggested by Riolo, Cohen, and Axelrod (2001). The authors show that cooperation can arise when individuals donate to others who are sufficiently similar to themselves on an arbitrary characteristic or "tag." In their model, computer-generated agents contribute to each other only if they are sufficiently similar on a bounded, continuous, and arbitrary scale. Using an evolutionary model with inheritable tags, Riolo et al. illustrated how a population of agents is rapidly able to establish a substantial degree of cooperation. Nevertheless the actual decision rule does not reflect an individual's rationale. More specifically, it is not based on an individual's selfish desire to maximize payoffs. If individuals are allowed to choose whether to cooperate or to defect, they encounter the basic PD game dilemma, which motivates them to defect. Roberts and Sheratt (2001) showed that when agents are allowed

[1] Hofstadter (1985) challenged the need for metaphysical explanations by asking the reader to consider solving an arithmetic problem which has already been solved by several other people who have sealed their solutions in envelopes. Having already decided what the solution is, a reader may recheck his calculations and occasionally find a mistake. By correcting his initial calculations, he is not assuming a change in the content of the envelopes but rather a change in his image of the content of the other envelopes.

to choose between cooperation and defection in the traditional manner, cooperation rates in Riolo et al.'s model decline to almost negligible percentages (1.48% vs. 73.6% in Riolo et al.'s model). These findings do not refute Riolo et al.'s model as a plausible mechanism for the evolution of cooperation among strangers. Similarity-based cooperation may indeed emerge as an arbitrary mutation that offers substantial fitness and is hence selected while competing with other, less advantageous behaviors. Alternatively, as proposed in the current article, cooperation may also arise from a rationale that compares the game's payoff structure with a behavioral factor, namely the perception of opponent's similarity.

The next section revises the continuous symmetry argument by proposing the subjective expected relative similarity (SERS) theory, which allows for deriving specific behavioral predictions for each PD payoff matrix. The proposed theory differs from its predecessors in three meaningful ways: (a) it addresses each player's subjective perception of similarity; (b) it treats the decision each player makes as a decision under uncertainty and hence calculates an expected value for each choice; (c) it suggests a relative decision rule that makes behavior conditional not only on the perception of similarity but also on the payoffs associated with each specific PD game.

## Subjective Expected Relative Similarity (SERS)

In line with the continuous symmetry argument, SERS defines the similarity of the players along a bounded probabilistic scale, assuming all values between 0 or complete dissimilarity and 1 or complete similarity. Hence, SERS rephrases the PD decision as a decision under uncertainty and suggests its solution by using decision-making-under-uncertainty models that assign preferences to alternatives by calculating the product of probability and payoff. (For example, the expected value of a lottery that pays $100 if a tossed coin falls on heads and $0 otherwise is 0.5 × $100 = $50.) These models have been developed from expected value (Huygens, 1657/1998) to subjective expected utility (Von Neuman & Morgenstern, 1947) and prospect theory (Kahneman & Tversky, 1979). For reasons of clarity and simplicity, I will use the simpler form of expected value. The modifications required to change from expected value to expected utility or prospect theory are not unique to the proposed model. They may easily be derived from the decision-making literature, and hence need not be specified in the current article.

SERS assumes that the similarity between the players is subjectively and individually perceived. Two players confronting each other may have either identical or different perceptions of their similarity to their opponent. In other words, similarity perceptions need neither be symmetric nor correspond to formal logic constraints. For example, Tversky (1977) showed that Americans may consider Taiwan to be more similar to China than China is similar to Taiwan. Formally, the similarity between the players is denoted by a subjective probability, $p_s$, expressing the degree of belief that one's opponent will act in a similar manner to oneself, where $0 \leq p_s \leq 1$ ($p_s = 1$ corresponds to complete similarity, $p_s = 0$ corresponds to complete dissimilarity, and $p_s = 0.5$ corresponds to the expectation that the chances the opponent will choose a similar or a dissimilar alternative are equally likely). After perceiving $p_s$, each player chooses between cooperation and defection, attempting to maximize the expected outcome. This means that each player estimates his or her expected payoffs under each of the possible courses of action, as shown in the decision tree depicted in Figure 1. The expected payoff of cooperation integrates the chances of obtaining the reward ($R$) payoff for mutual cooperation and the chances of obtaining the sucker's ($S$) payoff following unreciprocated cooperation. Because $p_s$ denotes the prospects of making a choice identical to that of the opponent (i.e., both players choose to cooperate or both players choose to defect), and $1 - p_s$ denotes the complementary prospects (i.e., choosing to cooperate while the opponent defects or choosing to defect while the opponent cooperates), the expected value of cooperation is given by $R \times p_s + S \times (1 - p_s)$. Similarly, the expected payoff of defection integrates the chances of obtaining the punishment ($P$) payoff for mutual defection and the chances of obtaining the temptation ($T$) payoff for unreciprocated defection and is given by $P \times p_s + T \times (1 - p_s)$. Ultimately a player has to compare both expected values and choose the course of action that promises a better outcome. Note that the player does not have a dominant strategy! The strategy that maximizes expected payoffs is defined as follows:

Cooperate if
$$R \times p_s + S \times (1 - p_s) > P \times p_s + T \times (1 - p_s)$$

Be indifferent (either cooperate or defect) if
$$R \times p_s + S \times (1 - p_s) = P \times p_s + T \times (1 - p_s)$$

Otherwise defect

This decision rule may also be expressed as:

$$\text{Cooperate if } p_s > \frac{T - S}{T - S + R - P}$$

$$\text{Be indifferent if } p_s = \frac{T - S}{T - S + R - P}$$

Otherwise defect

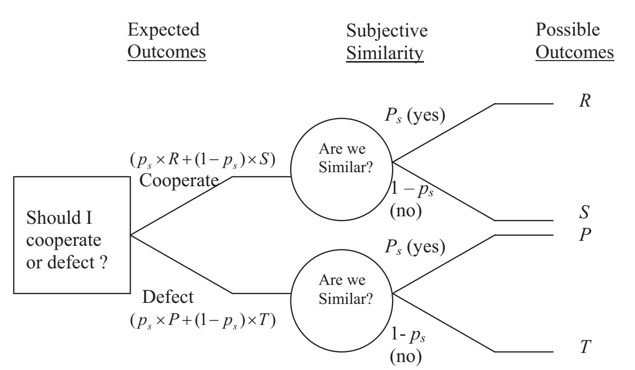


*Figure 1.* A decision tree with expected value calculations for the PD game, conditional upon the subjective probability assigned to the opponent's similarity. The payoffs are denoted as reward ($R$), sucker ($S$), punishment ($P$), and temptation ($T$); $p_s$ indicates the probability assigned to the similarity between player and opponent. To maximize expected payoffs, players should choose to cooperate for every $p_s > p_s^*$, where $p_s^* = \frac{T - S}{T - S + R - P}$ is the similarity threshold of the game.

Because the ratio $\frac{T-S}{T-S+R-P}$ is defined by the game's payoff matrix, it provides an objective criterion for comparison[2] with the subjective perception of similarity. Thus, defining the critical similarity of the game, or $p_s^*$, as $p_s^* = \frac{T-S}{T-S+R-P}$, we obtain a simple and relative payoff-maximizing decision rule: Cooperate if $p_s > p_s^*$, be indifferent if $p_s = p_s^*$, otherwise defect. Note that this rule incorporates both a behavioral component, namely the perception of similarity, and a situational component defined by the payoffs of the game. Each PD payoff matrix defines a specific $p_s^*$ threshold ($0.5 < p_s^* < 1$)[3] for comparison with the subjectively perceived similarity.

At this point, two possible misunderstandings need to be clarified. First, the relation between SERS and game theoretic analysis needs to be explained, for one may wonder whether the two contradict one another. For example, consider an individual applying the SERS principle and concluding that given a certain payoff matrix and his or her perception of the opponent's similarity, cooperation is indeed the preferred choice (because the perceived opponent's similarity exceeds the game's similarity threshold). But suppose now that this individual continues to reason that because this is also the course of action expected by the opponent (assuming that the opponent's subjective perception of similarity does not differ from his or her own perception), it should be possible to gain even more by exploiting the opponent and choosing to defect. The fundamental error in this line of reasoning is that a high level of similarity indicates a low probability of unilateral choices. If, on the one hand, the perception of similarity is high, the prospects of making unilateral choices (such as exploiting a cooperative opponent) are low. On the other hand, if the player assumes he or she is likely to make uncoordinated choices, the perceived similarity level with the opponent should reflect this perception and be assigned a low value. Indeed there are many reasons that may lead players to conclude they are capable of outperforming an opponent. Players may consider themselves more competent and experienced than they really are (Dunning, Johnson, Ehrlinger, & Kruger, 2003), luckier (Wagenaar & Keren, 1988), or in possession of a better memory and hence able to reason at least one step further than the opponent (Camerer, Ho, & Chong, 2004; Stahl & Wilson, 1995). In all these cases, the logical next step is to maximize the payoff by defecting! This conclusion is identical, both under game theory (which recommends defection for any single-step PD) and the proposed SERS extension (because it reflects the player's disbelief in the existence of a sufficient level of similarity). In other words, while SERS extends and complements game theoretic analysis, in many cases it recommends defection exactly as the regular game theoretic analysis does.

The second possible misunderstanding relates to a false attribution of causality. Because SERS relies on similarity estimates, it may bring to mind other similarity-related theories. More specifically, the literature has distinguished between three types of self-focused similarity estimates. It has been suggested that voting behavior can be explained by false accounts of causality where people behave as though acts that are merely correlated with a particular outcome actually cause that outcome (Quattrone & Tversky, 1984). Such beliefs can be classified as magical thinking or as examples of the illusion of control phenomenon (Langer & Roth, 1975). It has also been suggested that people regard their own actions as diagnostic, being aware of the fact that their choices do not cause those outcomes (Quattrone & Tversky, 1984). A third category, termed quasi-magical thinking, describes instances in which people act as if they erroneously believe their action influences the outcome, even though they do not really hold that belief (Shafir & Tversky, 1992). It is important to note that the similarity in SERS does not relate to any of the above. In contradiction to all three forms of self-focused similarity, SERS requires truly comparative similarity estimates. It assumes that individuals search for the best available indicators that allow comparing self and other and derive genuine similarity estimates. Nevertheless, this comparison builds upon uncertain cues with various validities. Some cues provide highly accurate information, whereas others may consist of anecdotal or misleading information.

To summarize, subjective expected relative similarity (SERS) embeds three meaningful features: (a) it assumes a continuous and uncertain range of similarities, (b) it allows subjective and thus not necessarily symmetric judgments of similarity, and (c) it is relative because it conditions an individual's decision upon the relation between the individual's subjective evaluation of the expected similarity ($p_s$) and the critical similarity index, defined by the payoffs ratio $\frac{T-S}{T+R-S-P}$ or $p_s^*$. Most importantly, SERS defines and predicts the conditions that should motivate payoff-maximizing players to cooperate while playing single-step PD games.

## SERS as a Descriptive Theory of Behavior

To consider SERS not only as a behavioral extension of the normative solution (because it requires the addition of a behavioral variable, namely the perception of the opponent's similarity) but also as a possible descriptive theory of genuine human behavior requires validating its behavioral predictions under various conditions. The described research program moves gradually from portraying a theoretical thought experiment to conducting empirical studies of genuine two-person interactions. Hence, the following sections (a) demonstrate the rationality of SERS following the detection of completely valid similarity cues, (b) validate SERS's prediction in genuine two-person interactions under the induction of experience-based similarity cues, and (c) validate SERS's prediction in genuine two-person interactions following the induction of semantic similarity cues.

## A Thought Experiment Illustrating SERS's Rationality

To picture a PD game that is played against an opponent whose similarity level can be portrayed by a precise probability, imagine a thought experiment where a genuine player (say you) plays against his or her own mirror reflection. The player is placed in

[2] Further transformations of values into subjective utilities (i.e., $U(x) = x^{1/2}$) may also reveal some individual differences.

[3] To obtain the lower boundary of $p_s^*$, while adhering to the PD constraints of $T > R > P > S$: let $R \rightarrow T$, and $P \rightarrow S$, and hence $\frac{T-S}{T-S+R-P} \rightarrow 0.5$. To obtain the higher boundary of $p_s^*$: let $(R-P) \rightarrow 0$, and hence $\frac{T-S}{T-S+R-P} \rightarrow 1$.

front of a mirror and is asked to draw a card from one of two boxes, a cooperation box containing a face-down cooperation card and a defection box containing a face-down defection card. The player is required to choose one of the boxes, pick up the card (without turning it over), and hold it in front of the mirror. The outcome of the game is determined first by the choice of box (either cooperation or defection) and second by the face value of the card reflected in the mirror (either cooperation or defection). If, as described, both cards are congruent with their boxes, a genuine player has to choose between mutual cooperation and mutual defection. If, however, the cards are non-congruent, so that the defection card is placed in the cooperation box and the cooperation card is placed in the defection box, a genuine player has to choose between being exploited (cooperating while playing against a defecting opponent) or exploiting the opponent (defecting while playing against a cooperating opponent).

Expanding this discrete card selection scenario into a decision under uncertainty yields a more interesting task. To this end, imagine that both the cooperation and the defection boxes are filled with 100 face-down cards. If some of the cards are congruent with the box they are drawn from and some are not, the mix of congruent and non-congruent cards allows simulating various similarity levels. For example, if the box contains 50 congruent cards and 50 non-congruent cards, similarity is given by $p_s = 0.5$. For each such proportion and critical similarity, $p_s^*$, there exists a decision, either defection or cooperation, that maximizes the player's expected payoff.

To demonstrate the player's best choices along the continuum of similarity levels, consider a PD game with known parameters, 100 congruent (cooperation) cards placed in the cooperation box and 100 congruent (defection) cards placed in the defection box. After the player makes the first payoff-maximizing choice (cooperation), five cooperation cards from the cooperation box are replaced with five defection cards from the defection box. The two decks are then reshuffled and placed back in their respective boxes. After the player's second choice, five additional congruent cards from the cooperation box are replaced with five additional congruent cards from the defection box, decreasing the level of similarity between the genuine player and his mirror reflection. After 20 replacements, both boxes have become completely incongruent with the cards placed in them. Clearly, rational players are expected to start the game with the choice of the cooperation box, and finish it with the choice of the defection box. This illustrates the fact that altering similarity levels changes conditional probabilities of cooperation and defection (i.e., cooperation of the mirror reflection given the cooperation of the genuine player) and underlines the importance of similarity perception in the SERS theory. Nevertheless, this thought experiment is an artificial scenario that completely overrides natural perception. Validating SERS as a genuine descriptive theory of actual human behavior requires conducting concrete experimental studies.

## Study 1: Testing SERS With Genuinely Experienced Similarity Cues

The first study aims to demonstrate the natural use of the SERS decision rule by randomly creating pairs of people who perceive each other as either similar or dissimilar and examining whether they differ in their levels of cooperation.

To this end the study applies a rather unusual paradigm that induces powerful perceptions of apparent similarity. The study uses a random process to create varying degrees of similarity in pairs of participants, which are then predicted to influence cooperation levels.

The study's hypotheses are that (a) participants assigned to low similarity threshold games (i.e., games with a relatively low $\frac{T-S}{T-S+R-P}$ ratio) will cooperate more often than participants assigned to high similarity threshold games (i.e., games with a relatively high $\frac{T-S}{T-S+R-P}$ ratio) and that (b) participants who perceive a high similarity level with their opponent will cooperate more often than participants who perceive a low similarity level with their opponent.

### Method

*Participants.* Four-hundred and fifty-nine participants studying toward a bachelor of arts in various disciplines participated in the experiment in return for performance-contingent payments.

*Apparatus.* This consisted of a table with an opaque screen affixed to its center, two pairs of cards showing either an ampersand (&) or a pound (#) symbol, two different PD payoff matrixes ($T = 20$, $R = 14$, $P = 2$, $S = 0$, corresponding to a $p_s^*$ value of 0.63 and $T = 20$, $R = 10$, $P = 5$, $S = 0$, corresponding to a $p_s^*$ value of 0.80), and written instructions for understanding and playing the PD matrix game.

*Procedure.* Three-hundred and ninety-four participants were randomly paired and seated on opposite sides of the opaque screen. They were then given an identical pair of two different cards (one card showing an "&" and the other a "#" symbol) and asked to coordinate their thoughts with those of the other participant, so that both select the same card. After both participants selected a card and placed it on the table, the screen was removed and the selected cards were revealed. Repeating the procedure three times results in four possible numbers of coordinated choices: 0, 1, 2, or 3, occurring with probabilities of 1/8, 3/8, 3/8, 1/8, respectively (assuming random and independent choices across trials and participants and applying the binomial distribution). After this arbitrary induction of similarity, participants were dismissed and asked to take part in another short experiment in a nearby cubicle. After being welcomed by the new experimenter, the pairs were shown one of the two PD matrixes and explained how to interpret its payoffs. Participants were randomly assigned to either a low similarity threshold matrix ($T = 20$, $R = 14$, $P = 2$, $S = 0$, and $p_s^* = 0.63$) or a high similarity threshold matrix ($T = 20$, $R = 10$, $P = 5$, $S = 0$, and $p_s^* = 0.80$). They were asked to play a single game by simultaneously writing down their choice on a sheet of paper and handing it over to the experimenter. Finally, both participants were paid in accordance with the game's outcome. Sixty-five additional participants were assigned to two control groups and played one of the two PD games ($p_s^* = 0.63$ and $p_s^* = 0.80$) without undergoing similarity induction first. For these participants, the first part of the experiment was described as a test of statistical norms. They were asked to choose one of two cards for three successive turns, yet they were not instructed to coordinate their thoughts and were not allowed to see each other's choices (the opaque screen was kept in place, preventing participants from seeing the choices made by their partners).

### Results and Conclusions

As predicted by the first hypothesis, participants who played the high similarity threshold game ($p_s^* = 0.80$) cooperated in 56% of the games, whereas those who played the low similarity threshold game ($p_s^* = 0.63$) cooperated in 75% of the games, $\chi^2(1) = 17.30$, $p < .05$. As predicted by the second hypothesis, the results reveal a monotonic positive relation between the number of coordinated cards (0 in the control group and 1, 2, and 3 in the experimental groups) and the proportion of cooperative choices (52, 65, 62, 70, and 74, respectively), $\chi^2(4) = 9.62$, $p < .05$. Figure 2 depicts cooperation rates under both similarity thresholds and the various similarity-induction conditions. The figure shows that cooperation rates under the low similarity threshold control condition comprise about 67% of the outcomes and gradually increase to over 80% following a series of three successful card coordination attempts. Participants assigned to the high similarity threshold condition cooperated only 40% of the time under the control condition and gradually increased the extent of cooperation to about 70%, with some exceptions for zero successful card coordination attempts.

Examination of the results of the two control groups alone reveals a difference of 27% in cooperation between the low and high similarity threshold games (0.63 and 0.80, respectively), $\chi^2(1) = 4.61$, $p < .05$. The results also indicate that both groups cooperated less frequently than did the experimental groups with zero successful coordination attempts. It seems that this difference may be attributed to the salience of coordination and similarity in the experimental conditions, and to some extent also to the perception of paradoxical similarity required to remain uncoordinated throughout all three trials. Nonetheless, the overall pattern supports both hypotheses, showing that lower similarity thresholds induce higher cooperation levels and that being more successful in the card coordination task predicts higher cooperation levels in the subsequent PD game.

Although these results provide clear support for SERS prediction, one may still speculate that the exciting experience created by the experiment evoked a sense of sharing a special adventure and that cooperation was to some extent due to the sentiment evoked by this unique experience. Endorsing such an explanation does not refute SERS because the perception of similarity is indeed the perception of shared characteristics, including a common history or joint experiences. Nevertheless, to generalize the results and further validate the role of the perception of relative similarity, the second study separates the players and uses a different similarity induction method.

## Study 2: Testing SERS With Semantically Induced Similarity Cues

The second study aims to replicate the results of Study 1 by providing the participants with semantic similarity cues and examining their subsequent cooperation in a PD game. To this end, participants are given the opportunity to compare their own ratings on an attitude questionnaire with those of an otherwise anonymous opponent. The comparison generates perceptions of either high or low similarity levels, which according to SERS are expected to influence the extent of cooperation or defection in a subsequent PD game.

### Method

*Participants.* Two-hundred and fifteen undergraduate students participated in the study in return for class credit or for a show-up fee for filling in the questionnaire as well as for a performance-contingent payment, depending on the outcome of a single PD game.

*Apparatus.* First, a 7-point scale attitude questionnaire was administered, consisting of the following eight items:[4] (a) I am interested in social and interpersonal issues; (b) My thoughts are deeper and more profound than they seem from the outside; (c) I am interested in technical and engineering topics; (d) Sometimes I don't need words to understand the intentions of others; (e) I often act without considering the outcomes of my actions; (f) My interpersonal relations are more complex than they appear at first glance; (g) Every so often I have deep and surprising insights; and (h) Sometimes I take risks and sometimes I am cautious and careful. Note that the second, seventh, and eighth items address the individual's self-esteem and induce a relatively broad consensus. Others, such as the first and third items, produce consensus among specific population groups, such as psychology or engineering students, respectively. The remaining three items are rather personal and do not have a commonly agreed-upon truth value. Second, there were two PD payoff matrixes ($T = 20$, $R = 10$, $P = 5$, $S = 0$, $p_s^* = 0.63$, and $T = 20$, $R = 14$, $P = 2$, $S = 0$, $p_s^* = 0.80$) and written instructions for understanding and playing the PD matrix game.

*Procedure.* The study comprised two sessions. In the first session participants were asked to complete the eight-item attitude questionnaire. They were then individually invited to participate in an experiment at the laboratory about two days later. During the second session each participant received his or her own previously rated questionnaire together with that of an otherwise anonymous participant. In fact, participants were assigned either to a similar or a dissimilar condition and were shown fabricated opponent questionnaires. The similar questionnaires comprised attitude ratings that deviated by 0, 1, or 2 points (with an equal random distribution) from the actual participant's response, and the dissimilar questionnaires comprised attitude ratings that deviated either by 3 or by 4 points (with an equal random distribution). It was explained that the identity of all participants was being kept secret, but that each could still learn something about the other by comparing the questionnaire ratings. Further, participants were asked to compose a short written paragraph addressing the attitudes expressed in both questionnaires and to conclude by checking one of three boxes: *quite similar*, *moderately similar*, or *not similar*. This was in order to ensure that participants paid sufficient attention to all ratings as well as to check the similarity manipulation.

[4] The items of the attitude questionnaire were selected on the basis of a pilot study involving 20 participants who were asked to rate 17 initial items addressing personal attitudes and preferences. The questionnaire avoided issues that relate to political disputes and might indicate conflict-related opinions or strategies. Based on this initial screening, eight items that generated strong responses (rated either low or high) were selected for inclusion in the actual study.

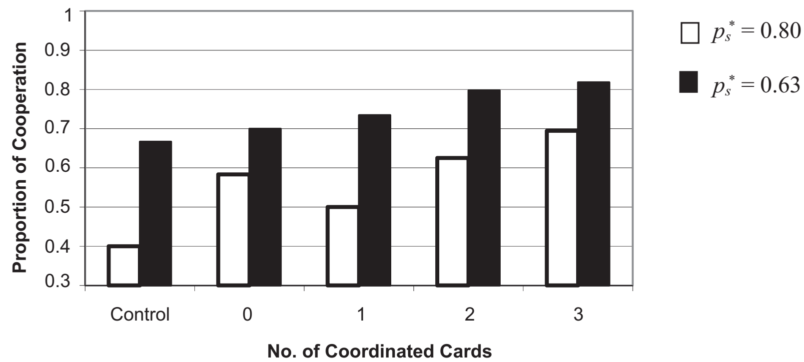


*Figure 2.* The proportion of cooperation choices as a function of the number of successful card coordination attempts (none in the control conditions and 0, 1, 2, or 3 in the experimental conditions) and two prisoner's dilemma matrixes with different similarity thresholds ($p_s^* = 0.63$ and $p_s^* = 0.80$).

Although this procedure prompts participants to pay attention to the extent of similarity, it does not provide clues to the study's actual hypotheses. Participants could of course speculate that the experimenter was interested in the link between similarity and behavior, but there was nothing in the situation to suggest what response was expected, given a subjectively perceived similarity level and a specific payoff matrix.

After participants completed the questionnaire examination, the game matrix was shown and explained. Next, each participant was asked to make a choice, mark it on a sheet of paper, and seal it in an envelope identified only by a secret code. The participants were told that after the entire experiment was completed, the experimenter would open all the envelopes, determine the outcomes, and pay each participant the deserved payoff. Because the opponent questionnaires had actually been prepared by the experimenters between the two sessions, the outcomes were determined by matching participants from the same experimental group (reflecting the actual similarity perceptions experienced by the participants). This procedure allowed generating fair payoffs, but it had no impact on the actual experimental results.

### *Results and Conclusions*

As predicted by the first hypothesis, participants playing the low similarity threshold game cooperated more often than those playing the high similarity threshold game, $\chi^2(1) = 6.52$, $p < .05$. Cooperation rates were 30% under the high similarity threshold condition ($p_s^* = 0.80$) and 46% under the low similarity threshold condition ($p_s^* = 0.63$). As predicted by the second hypothesis, participants who believed their opponent responded in a similar manner cooperated more often than participants who believed their opponent responded in a dissimilar manner, $\chi^2(1) = 7.10$, $p < .05$. Participants cooperated 47% of the time under the similar condition and 29% of the time under the dissimilar condition. Table 2 describes the fourfold pattern of cooperation rates under the two levels of similarity perception and the two similarity thresholds.

## General Discussion

The current article has examined the perception of similarity as a driving behavioral mechanism in PD games that involve both the motivation to cooperate and the temptation to defect. A new theory of subjective expected relative similarity (SERS) has been developed based on the problematic symmetry argument and the improved continuous and uncertain version of the argument. This theory suggests that similarity is subjectively perceived along a bounded continuum and that this perception is compared with a situational similarity threshold derived from the possible payoffs. It was hypothesized that games with lower similarity thresholds induce more cooperation (because individuals are more likely to encounter sufficiently similar opponents) and that opponents who are perceived as being more similar than the similarity threshold of the game will motivate cooperative behavior. Notwithstanding the need for further validation of the results, the obtained data show that SERS is indeed a descriptive theory with the power to predict the extent of cooperation and defection in single-step PD games.

What, then, is the wider significance of SERS? Five basic models have been shown to account for the evolution of cooperation: direct reciprocity, indirect reciprocity, network reciprocity, kin selection, and group selection (Nowak, 2006). While reciprocity models are dependent on various formulations of repeated interactions, both kin and group selection models depend on the ability to identify either genetic or social relatedness. For cooperation to emerge and stabilize under these mechanisms, organisms need to make cooperative behavior contingent on the probability of either genetic or social relatedness. Defining $p_r$ as an uncertain or probabilistic estimate of relatedness, and $\alpha$, $\beta$, $\gamma$, and $\delta$ as the payoffs respectively associated with cooperation following correct identification of relatedness, cooperation following misidentified relatedness, defection following false rejection of relat-

Table 2
*Cooperation Percentages Under Two Levels of Similarity Perception and Two Similarity Thresholds*

| Similarity threshold | Experimental condition: Similar | Dissimilar |
|---|---|---|
| 0.63 | 59% | 33% |
| 0.80 | 34% | 25% |

edness, and defection following correct rejection of relatedness, we obtain a decision rule that recommends cooperation whenever $\alpha \times p_r + \beta \times (1 - p_r) > \gamma \times p_r + \delta \times (1 - p_r)$ or whenever $p_r > \frac{\delta - \beta}{\delta - \beta + \alpha - \gamma}$.

Note, however, that $p_r$ is a true and concealed state of nature. How can one estimate $p_r$? Although family members may recognize their relatives and members of small communities may easily distinguish a member of an in-group from that of an out-group, constituents of larger populations need to rely on proxies for estimating either genetic or social relatedness. Human individuals may use cues such as physical appearance, clothing, jewelry, language, accent, rituals, or emblems to distinguish between in- and out-group members. In fact, research on the "minimal group paradigm" has shown that trivial performance criteria (Tajfel, 1970, 1974; Tajfel & Turner, 1979) and dependence upon the allocations of others (Gaertner & Insko, 2000) are sufficient conditions for the emergence of group identity that increases both in-group cooperation and out-group discrimination. Substituting the true yet concealed $p_r$ with the perceived $p_s$ and defining the interaction as a PD game by replacing α, β, γ, and δ with the *R*, *S*, *P*, and *T* payoffs, respectively, we obtain the SERS solution. The empirical results and analyses suggest that (a) both cooperative and competitive behaviors in single PD games coincide with sound individual rationality and that (b) SERS is a likely source for the emergence of both kin and group selection behaviors in large populations. Nevertheless, these conclusions depend on the abilities to detect, process, and apply similarity. In fact, these abilities are known to function even in earlier stages of phylogenesis. It has been shown that mimicry of color, movement, and shape provides powerful protection against predators and also helps in luring prey (Wickler, 1968). In some cases mimicry enables parasites to elicit cooperative behavior from their hosts. One example is the common cuckoo or *Cuculus canorus*, which has evolved so that its eggs mimic the color, size, and shape of warbler *Acrocephalus arundinaceus* eggs, misleading the warbler female and causing her to incubate the cuckoo's eggs (Payne, 1997). In humans, similarity and mimicry have been shown to induce interpersonal attraction (Byrne, 1971), affiliation (Lakin & Chartrand, 2003), rapport (Chartrand & Bargh, 1999), and generosity (Van Baaren, Holland, Steenaerts, & Van Knippenberg, 2003). These phenomena and behaviors all point to the central role of similarity perception in survival and interaction with other individuals.

Still, there is a significant gap between the ability to perceive similarity and its useful implementation in mixed-motives interactions such as the single-step PD examined in the current research. Clearly, neither the participants in Studies 1 and 2 nor naturally behaving individuals make formal computations or knowingly compare similarity perceptions with similarity indices.

For SERS to function in natural settings, individuals need to have evolved a link between similarity perception and the arousal of either positive or negative emotions. Such a mechanism, revealing a direct relation between attitude similarity and interpersonal attraction, has indeed been studied and described by the "attitude similarity–attraction" paradigm (Byrne, 1971). Combining SERS with this paradigm suggests a multifaceted process involving cognition, emotion, and behavior. It indicates that individuals with similar attitudes are expected to like each other and engage in cooperative interactions, whereas individuals that differ with respect to meaningful attitudes are expected to feel rejection and engage in hostile interactions. According to Young (1971), potent attitudes that generate strong emotions are related to attitudes about God, war, college education, the American way of life, and premarital sexual relations. Notwithstanding the impact of societal changes over the past four decades, these issues still generate public agendas and political disputes.

Finally, I would like to conclude by noting that SERS contributes to a vast and growing research body examining the evolution of cooperation. Although game theoretic analysis of single-step PD games has a clear and dominant normative solution in the form of defection, evolution seems to have generated cooperative and prosperous societies capable of overcoming this barrier by developing social norms (Fehr & Fischbacher, 2003; Gintis et al., 2003), learning from individual experiences (Erev & Roth, 1998, 2002), observing others (Chater, Vlaev, & Grinberg, 2008; Fischer, 2003a, 2003b), and, as shown in the current research, relying on the perception of relative similarity.

## References


Axelrod, R. (1980a). Effective choice in the prisoner's dilemma. *Journal of Conflict Resolution, 24,* 3–25.

Axelrod, R. (1980b). More effective choice in the prisoner's dilemma. *Journal of Conflict Resolution, 24,* 379–403.

Axelrod, R. (1981). The emergence of cooperation among egoists. *The American Political Science Review, 75,* 306–318.

Axelrod, R. (1984). The success of TIT FOR TAT in computer tournaments. In *The evolution of cooperation* (pp. 27–54). New York: Basic Books.

Axelrod, R., & Hamilton, W. D. (1981, March 27). The evolution of cooperation. *Science, 211,* 1390–1396.

Bornstein, G. (2004). Cooperation in intergroup and single-group social dilemmas. In R. Suleiman, D. V. Budescu, I. Fischer, & D. Messick (Eds.), *Contemporary psychological research on social dilemmas* (pp. 227–247). Cambridge, United Kingdom: Cambridge University Press.

Brams, S. (1975). *Game theory and politics.* New York: Free Press.

Byrne, D. (1971). *The attraction paradigm.* New York: Academic Press.

Camerer, C. F., Ho, H., & Chong, J. (2004). A cognitive hierarchy model of games. *Quarterly Journal of Economics, 119*(3), 861–898.

Campbell, R. (1985). Background for the uninitiated. In R. Campbell & L. Sowden (Eds.), *Paradoxes of rationality and cooperation: Prisoner's dilemma and Newcomb's problem* (pp. 3–41). Vancouver: University of British Columbia Press.

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception–behavior link and social interaction. *Journal of Personality and Social Psychology, 76,* 893–910.

Chater, N., Vlaev, I., & Grinberg, M. (2008). A new consequence of Simpson's paradox: Stable cooperation in one-shot prisoner's dilemma from populations of individualistic learners. *Journal of Experimental Psychology: General, 137,* 403–421.

Danielson, P. (1992). *Artificial morality: Virtuous robots for virtual games.* New York: Routledge.

Davis, L. H. (1985). Prisoners, paradox, and rationality. In R. Campbell & L. Sowden (Eds.), *Paradoxes of rationality and cooperation: Prisoner's dilemma and Newcomb's problem* (pp. 43–49). Vancouver: University of British Columbia Press.

Dawes, R. M. (1980). Social dilemmas. *Annual Review of Psychology, 31,* 169–193.

Dixit, A. K., & Nalebuff, B. J. (1992). *Thinking strategically.* New York: Norton.

Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why people

fail to recognize their own incompetence. *Current Directions in Psychological Science, 12*(3), 83–87.
Erev, I., & Roth, A. E. (1998). Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *American Economic Review, 88,* 848–881.
Erev, I., & Roth, A. E. (2002). Simple reinforcement learning models and reciprocation in the prisoner's dilemma game. In G. Gigerenzer & R. Selten (Eds.), *Bounded rationality: The adaptive toolbox* (pp. 215–231). Cambridge, MA: MIT Press.
Fehr, E., & Fischbacher, U. (2003, October 23). The nature of human altruism. *Nature, 425,* 785–791.
Fehr, E., & Gaechter, S. (2002, January 10). Altruistic punishment in humans. *Nature, 415,* 137–140.
Fischer, I. (2003a). The emergence of reactive strategies in simulated heterogeneous populations. *Theory and Decision, 55*(4), 289–314.
Fischer, I. (2003b). Evolutionary development and learning: Two facets of strategy generation. *Journal of Artificial Societies and Social Simulations, 6*(1). Available from http://jasss.soc.surrey.ac.uk/6/1/7.html
Fischer, I., & Suleiman, R. (1997). Mutual cooperation in a simulated inter-group conflict. *Journal of Conflict Resolution, 41*(4), 483–508.
Flood, M., & Dresher, M. (1952). Some experimental games. *Research memorandum RM-789.* Santa Monica, CA: Rand.
Frank, R. H., Gilovich, T., & Regan, D. T. (1993). The evolution of one-shot cooperation: An experiment. *Ethology and Sociobiology, 14,* 247–256.
Gaertner, L., & Insko, C. A. (2000). Intergroup discrimination in the minimal group paradigm: Categorization, reciprocation, or fear? *Journal of Personality and Social Psychology, 79,* 77–94.
Gauthier, D. (1985). Maximization constrained: The rationality of cooperation. In R. Campbell & L. Sowden (Eds.), *Paradoxes of rationality and cooperation: Prisoner's dilemma and Newcomb's problem* (pp. 75–93). Vancouver: University of British Columbia Press.
Gintis, H., Bowles, S., Boyd, R., & Fehr, E. (2003). Explaining altruistic behavior in humans. *Evolution and Human Behavior, 24,* 153–172.
Hardin, G. (1968, December 13). The tragedy of the commons. *Science, 162,* 1243–1248.
Hofstadter, D. R. (1985). *Metamagical themas: Questing for the essence of mind and* pattern. New York: Basic Books.
Howard, J. V. (1988). Cooperation in the prisoner's dilemma. *Theory and Decision, 24,* 203–213.
Huygens, C. (1998). Du calcul dans les jeux de hasard [Of the calculus in games of chance]. In *Oeuvres complètes* (Vol. 14). The Hague, The Netherlands: Martinus Nijhoff. (Original work *Ratiociniis in Aleae Ludo,* published 1657)
Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47,* 263–291.
Lakin, J. L., & Chartrand, T. L. (2003). Using nonconscious behavioral mimicry to create affiliation and rapport. *Psychological Science, 14,* 334–339.
Langer, E. J., & Roth, J. (1975). Heads I win, tails it's chance: The illusion of control as a function of the sequence of outcomes in a purely chance task. *Journal of Personality and Social Psychology, 32,* 951–955.
Laplace, M. (1953). *Essay on probabilities.* Mineola, NY: Dover. (Original work published 1783)
Lewis, D. (1979). Prisoners' dilemma is a Newcomb problem. *Philosophy and Public Affairs, 8*(3), 235–240.
Luce, D. R., & Raiffa, H. (1957). *Games and decisions.* New York: Wiley.
Macy, M. W. (1990). Learning theory and the logic of critical mass. *American Sociological Review, 55,* 909–926.
Macy, M. W. (1991a). Chains of cooperation: Threshold effects in collective action. *American Sociological Review, 56,* 730–747.
Macy, M. W. (1991b). Learning to cooperate: Stochastic and tacit collusion in social exchange. *American Journal of Sociology, 97*(3), 808–843.
McCabe, K. A., Rassenti, S. J., & Smith, V. L. (1996). Game theory and reciprocity in some extensive form experimental games. *Proceedings of the National Academy of Sciences, USA, 93*(13), 421–428.
Nowak, A. M. (2006, December 8). Five rules for the evolution of cooperation. *Science, 314*(8), 1560–1563.
Payne, R. B. (1997). Avian brook parasitism. In D. H. Clayton & J. Moore (Eds.), *Host–parasite evolution* (pp. 338–369). Oxford, United Kingdom: Oxford University Press.
Quattrone, G. A., & Tversky, A. (1984). Causal versus diagnostic contingencies: On self-deception and on the voter's illusion. *Journal of Personality and Social Psychology, 46,* 237–248.
Rapoport, A. (1960). *Fights, games, and debates.* Ann Arbor: University of Michigan.
Rapoport, A., & Chammah, M. (1965). *Prisoner's dilemma: A study in conflict and cooperation.* Ann Arbor, MI: Ann Arbor Press.
Rapoport, A., & Guyer, M. (1966). A taxonomy of $2 \times 2$ games. *General Systems, 11,* 203–214.
Riolo, R. L., Cohen, M. D., & Axelrod, R. (2001, November 22). Evolution of cooperation without reciprocity. *Nature, 414,* 441–443.
Roberts, G., & Sheratt, T. N. (2001, August 1). Does similarity breed cooperation? *Nature, 418,* 499–500.
Sally, D. (1995). Conversation and cooperation in social dilemmas: A meta-analysis of experiments from 1958 to 1992. *Rationality and Society, 7*(1), 58–92.
Schelling, T. C. (1963). *The strategy of conflict.* New York: Oxford University Press.
Schuessler, R. (1990). Threshold effects and the decline of cooperation. *Journal of Conflict Resolution, 34*(3), 476–494.
Sen, A. (1974). Choice, orderings and morality. In S. Korner (Ed.), *Practical reason* (pp. 54–67). New Haven, CT: Yale University Press.
Shafir, E., & Tversky, A. (1992). Thinking through uncertainty: Nonconsequential reasoning and choice. *Cognitive Psychology, 24,* 449–474.
Shubik, M. (1970). Game theory, behavior, and the paradox of the prisoner's dilemma: Three solutions. *Journal of Conflict Resolution, 14,* 181–193.
Stahl, D. O., & Wilson, P. W. (1995). On players' models of other players: Theory and experimental evidence. *Games and Economic Behavior, 10,* 218–254.
Tajfel, H. (1970). Experiments in intergroup discrimination. *Scientific American, 223,* 96–102.
Tajfel, H. (1974). Social identity and intergroup behavior. *Social Science Information, 13,* 65–93.
Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33–47). Monterey, CA: Brooks/Cole.
Tversky, A. (1977). Features of similarity. *Psychological Review, 84,* 327–352.
Van Baaren, R. B., Holland, R. W., Steenaerts, B., & Van Knippenberg, A. (2003). *Journal of Experimental Social Psychology, 39,* 393–398.
Von Neuman, J., & Morgenstern, O. (1947). *Theory of games and economic behavior.* Princeton, NJ: Princeton University Press.
Wagenaar, W. A., & Keren, G. B. (1988). Chance and luck are not the same. *Journal of Behavioral Decision Making, 1,* 65–75.
Wickler, W. (1968). *Mimicry in plants and animals.* London: World University Library.
Young, B. (1971). Analytic research on the base relationship. In D. Byrne (Ed.), *The attraction paradigm* (pp. 59–93). New York: Academic Press.

Received August 9, 2008
Revision received March 17, 2009
Accepted March 18, 2009 ■