Cognitive Constraints on How Economic Rewards Affect Cooperation

Ellen E. Furlong and John E. Opfer

The Ohio State University

ABSTRACT—Cooperation often fails to spread in proportion to its potential benefits. This phenomenon is captured by prisoner’s dilemma games, in which cooperation rates appear to be determined by the distinctive structure of economic incentives (e.g., $3 for mutual cooperation vs. $5 for unilateral defection). Rather than comparing economic values of cooperating versus not ($3 vs. $5), we tested the hypothesis that players simply compare numeric values (3 vs. 5), such that subjective numbers (mental magnitudes) are logarithmically scaled. Supporting our hypothesis, increasing only numeric values of rewards (from $3 to 300$) increased cooperation (Study 1), whereas increasing economic values increased cooperation only when there were also numeric increases (Study 2). Thus, changing rewards from 3¢ to 300¢ increased cooperation rates, but an economically identical change from 3¢ to $3 elicited no gains. Finally, logarithmically scaled reward values predicted 97% of variation in cooperation, whereas the face value of economic rewards predicted none. We conclude that representations of numeric value constrain how economic rewards affect cooperation.
THE PRESENT STUDIES

To examine how representations of numeric value influence the effect of economic rewards on cooperative behavior, we manipulated numeric value, both independently of economic value (Studies 1 and 2) and in combination with it (Studies 2 and 3), and observed four indices of IPD strategies: individual cooperation, mutual cooperation, mutual defection, and forgiveness.

The IPD is defined by relations between payoffs two players earn by cooperating or defecting (Fig. 1). This structure creates a dilemma in which individuals do best on any given iteration by defecting, yet overall both earn most by cooperating (Axelrod & Hamilton, 1981). Specifically, the reward for unilateral defection (T) is greater than the reward for mutual cooperation (R), which is greater than the reward for mutual defection (P), which is in turn greater than the reward for unilateral cooperation (S; see Fig. 1).

Thus, the reward structure present in the IPD—in which rewards for unilateral defection are greater than rewards for mutual cooperation (i.e., when R/T < 1; Rapoport & Chammah, 1965)—can explain irrationally low rates of cooperation.

Against this classical model, we hypothesized that cooperation depends on numeric structure of rewards and that manipulating only numeric values of R and T would affect cooperation in the IPD. That is, because payoffs for cooperating versus defecting are compared by brains representing numeric values logarithmically (Dehaene, 1997), and because logarithmic coding fails to preserve ratio information (Stevens, 1961), we expected that increasing numeric values of payoffs would make them less discriminable, thereby reducing players’ temptation to defect.

We tested our hypothesis by examining changes in cooperative behavior when numeric value increased but economic value was held constant ($3 to 300¢; Study 1), as well as when both numeric and economic values increased (3¢ to 300¢ and $3 to $300; Study 2). The linear model predicts no change in cooperative behavior with manipulations of numeric magnitude (e.g., $3 vs. 300¢); our model, however, predicts more cooperation for numerically larger rewards (300) than for numerically smaller rewards (3), regardless of economic value (3¢, $3, $300). To directly test our underlying theory, Study 3 examined cooperation under five conditions varying numeric and economic value over several orders of magnitude. We predicted that cooperation would be better predicted by ratios of logarithmically compressed numeric values—ln(R)/ln(T)—than by ratios of uncompressed values (R/T).

STUDY 1: TEMPTATION OF $3 VERSUS 300¢

Method

Thirty-one pairs of undergraduates were randomly assigned to one of two economically equivalent payoff matrices, one earning dollars (R = $3; S = $0; T = $5; P = $1) and one earning cents (R = 300¢; S = 0¢; T = 500¢; P = 100¢). Pairs were initially separated; one was chosen as Subject and one as confederate. Confederates played “Tit-for-Tat” (TFT), initially cooperating and thereafter copying the Subject’s behavior on the preceding trial. Subjects received no instruction on strategy but were introduced to payoff matrices and practiced 10 IPD trials with the Experimenter before playing the confederate; practice trials were not analyzed and served to introduce Subjects to procedures. Pairs were instructed to maximize earnings, posted after each of 80 trials.

Results and Discussion

Although dollars and cents conditions presented equivalent economic rewards, the cents condition elicited more individual cooperation, F(1, 21) = 6.90, p<.05, and mutual cooperation, F(1, 21) = 5.40, p<.05, was .25, and mutual defection, F(1, 21) = 9.21, p<.05, was .31, than the dollars condition did. Similarly, the dollars condition elicited greater mutual defection, F(1, 21) = 9.21, p<.05, was .97, and a longer latency to “forgive” the confederate, or to cooperate after the confederate’s first defection, F(1, 21) = 4.68, p<.05, was .89, .18, than the cents condition did (Fig. 2).

STUDY 2: EFFECT OF NUMERIC VERSUS ECONOMIC VALUE ON COOPERATION

To test whether higher cooperation rates for 300¢ rewards than for $3 rewards resulted from numeric values of rewards (300 vs. 3) rather than from a preference for dollars or cents, Study 2 presented subjects with numerically equivalent payoffs of both dollars and cents (3¢, 3¢; 300¢, 300¢). Additionally, subjects in Study 2 played a computer, thereby removing social feedback.
unit, four indices of player strategy revealed no main effect of multivariate analysis of variance (MANOVA) preference for dollars or cents, a 2 (units: dollars, cents) against the hypothesis that the results of Study 1 were due to a economic value, an F(1, 46) = 6.25, prep = .05, nor did unit interact with number, an F(4, 41) = .23, pprep = .16.

We next examined the effect of numeric value (3 or 300) and economic value ($0.03, $3, $300) for each of the four indices. Numerically greater rewards increased individual cooperation (3¢ versus .03, $3, $300 versus .01; mutual cooperation: $3 and $300 versus .001; mutual defection: $3 = .11 versus .05. However, the effect of partner type, $3 = .08, was greater than the effect of number, $3 = .05, for forgiveness, the only variable for which we observed a Partner x Number interaction, an F(1, 67) = 11.21, pprep = .99, $3 = .08.

**STUDY 3: EVIDENCE FOR LOGARITHMIC SCALING OF PAYOFFS**

In Studies 1 and 2, increasing numeric magnitudes increased cooperation, contradicting the critical assumption that behavior remains invariant when payoffs are transformed linearly (Rapoport & Chammah, 1965).
One way to understand this numeric-magnitude effect is to assume that numbers associated with payoff values are represented logarithmically, resulting in failure to conserve ratio information over linear transformations. That is, the linear model predicts defection whenever \( R/T \) is less than 1, and because 300¢/500¢ equals $3/$5, changing numeric values would not matter. This preservation of ratio information does not obtain if numeric values are scaled logarithmically, as \( \ln(300)/\ln(500) \) is approximately 1 (i.e., temptation to defect and cooperate are nearly equal), whereas \( \ln(3)/\ln(5) \) is approximately .68 (i.e., temptation to defect is higher than temptation to cooperate). Thus, logarithmic representations of numeric magnitude could explain numeric-magnitude effects in Studies 1 and 2.

To test quantitative predictions of this hypothesis, in Study 3 we generated new payoff matrices spanning several orders of magnitude by adding or multiplying a constant to all payoff values. The linear model predicts that multiplying constants will not change cooperation but that adding constants will increase cooperation. In contrast, the logarithmic model predicts that both manipulations will increase cooperation (Table 1).

### Method

Ninety-six undergraduates participated in Study 3. Procedures were identical to those in Study 1 except that, in addition to the baseline condition, a constant amount was added to (+100, +1,000) or multiplied by \((\times 0.001 \text{ or } \times 0.01)\) all baseline values \((R = 3¢, S = 0¢, T = 5¢, P = 1¢)\), resulting in five between-subjects conditions (Table 1).

### Results and Discussion

To test our logarithmic model, we regressed the four indices over the five conditions against two predictors: \( R/T \) and \( \ln(R)/\ln(T) \). The linear model, \( R/T \), accounted for virtually all the variance, \( R^2 = .97 \) (Fig. 4). The logarithmic model also accounted for more variance than did the linear model in rates of mutual cooperation, \( \ln(R)/\ln(T) \): \( R^2 = .71; \quad R/T: \ R^2 = .07; \quad \text{mutual defection, } \ln(R)/\ln(T): \ R^2 = .55; \quad R/T: \ R^2 = 0; \quad \text{and forgiveness, } \ln(R)/\ln(T): \ R^2 = .42; \quad R/T: \ R^2 = .02. \)

### GENERAL DISCUSSION

Cooperation often fails to spread in proportion to its potential benefits. This phenomenon is captured by IPD games, in which low cooperation rates appear to result from distinctive economic incentives \((T > R > P > S; \text{ Rapoport & Chammah, 1965})\). Thus, when confronted with larger rewards for unilateral defection \((S)\) than for mutual cooperation \((P)\) on a trial in an IPD, players often choose the economically larger reward, suggesting that linear transformations of monetary value \((e.g., \text{ converting }$5 \text{ to } 500¢)\) would not change cooperation \((\text{ Rapoport & Chammah, 1965})\). Challenging this assumption, we hypothesized that decisions involve comparing numeric rather than economic value and that, because mental representations of numeric value increase logarithmically, linear transformations of just numeric values would dramatically change cooperation rates.

Evidence for effects of numeric value on cooperation first came from manipulating numeric value while holding economic value constant \((\text{ Study 1})\), where rewards expressed in large numbers \((e.g., 300¢)\) elicited greater cooperation, less competition, and a shorter latency to forgive than did those expressed in small numbers \((e.g., 3¢)\), despite their monetary equivalence. Thus, a cooperation-to-defection ratio of 300¢:500¢ presented less temptation than a ratio of 3¢:5¢, an inequality following directly from logarithmic scaling of numbers — that is, \( \ln(300)/\ln(500) > \ln(3)/\ln(5) \). Further evidence came from Study 2, where changes in payoffs from 3¢ to 300¢ led to increased cooperation but an economically equivalent change from 3¢ to $3 did not. Thus, a cooperation-to-defection ratio of 3¢:5¢ was as tempting as $3:5¢, but both presented more temptation than the ratios 300¢:500¢ and $300:500, which elicited equal cooperation rates. Finally, as a quantitative test of our theory, we manipulated both reward value and numeric magnitude \((\text{ Study 3})\) and found that ratios of logarithmically compressed payoffs accounted for more variation in cooperative behavior than ratios of uncompressed payoffs did.

Effects of manipulating number over economically equivalent payoff matrices cannot be explained by preferences for dollar-versus penny-denominated rewards. In Study 2, changing only units \((i.e., 1¢ \text{ to } 1\text{¢})\) had no impact on cooperation. Nor were effects of number isolated to social situations: Numeric effects

### TABLE 1

The Linear and Logarithmic Models’ Predictions for the Five Matrices Used in Study 3

<table>
<thead>
<tr>
<th>Condition and matrix</th>
<th>Linear model: ( R/T )</th>
<th>Logarithmic model: ( \ln(R)/\ln(T) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;1&quot; condition: ( R = 3, \ S = 0, \ T = 5, \ P = 1 )</td>
<td>0.6</td>
<td>0.68</td>
</tr>
<tr>
<td>&quot;0.001&quot; condition: ( R = 0.003, \ S = 0, \ T = 0.005, \ P = 0.001 )</td>
<td>0.6</td>
<td>1.10</td>
</tr>
<tr>
<td>&quot;0.1&quot; condition: ( R = 0.03, \ S = 0, \ T = 0.05, \ P = 0.01 )</td>
<td>0.6</td>
<td>1.17</td>
</tr>
<tr>
<td>&quot;101&quot; condition: ( R = 103, \ S = 100, \ T = 105, \ P = 101 )</td>
<td>0.98</td>
<td>1.0</td>
</tr>
<tr>
<td>&quot;1,001&quot; condition: ( R = 1,003, \ S = 1,000, \ T = 1,005, \ P = 1,001 )</td>
<td>0.99</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Note. \( R \) = reward for mutual cooperation; \( T \) = temptation to defect; \( S \) = sucker’s reward; \( P \) = punishment for mutual defection.
were present whether the opponent was a human (Study 1) or computer player (Study 2). Indeed, number had a larger effect on individual cooperation, mutual cooperation, and mutual defection than did the effect of human participation.

Our findings fit into a wider literature examining logarithmic scaling of numeric magnitude. Across a range of numeric tasks (estimation and comparison of numeric magnitudes), age groups (infants, children, and time-pressured adults), and species (pigeons, rats, nonhuman primates, and humans), representations of numeric magnitude follow Fechner’s law, with differences between small quantities being overestimated and differences between large quantities being underestimated (Dehaene, 2007). One reason for activation of a logarithmic, analog magnitude system in the context of IPD games is that they require comparison of values expressed in Arabic numerals, which activate magnitude representations automatically—even when those representations interfere with task performance (Henik & Tselgov, 1982).

Logarithmic scaling is not a new characterization of numeric-magnitude representations, nor is it a new characterization of how monetary value affects decision making. Bernoulli’s (1738/1954) observation, “A gain of one thousand ducats is more significant to a pauper than to a rich man though both gain the same amount” (p. 24), set the stage for the value function in prospect theory, in which representations of monetary value were thought to follow the psychophysical regularity of Fechner’s law (Kahneman & Tversky, 1979, p. 278). We suggest that, instead of monetary value, it is representations of numeric value that are subject to this psychophysical regularity. Were monetary value rather than numeric value subject to Fechner’s law, changing rewards from 3¢ to 300¢ would elicit the same behavioral change as a change from 3¢ to $3—a hypothesis contradicted by our findings.

The suggestion that numeric rather than monetary value is scaled logarithmically may generalize to other economic-decision-making tasks involving numeric comparisons, including temporal discounting, bargaining, gambling, medical and insurance decisions, behavioral traps, and morality dilemmas. Finally, our findings suggest a novel explanation for the observation that economic games disproportionately activate the posterior parietal cortex. This observation has been variously explained by the somatosensory experience of reward and punishment (Bechara et al., 2005), attention to spatial locations (Colby, 1996), and the parietal cortex being an “economics module” (Camerer, Loewenstein, & Prelic, 2005; Glimcher, Dorris, & Bayer, 2005). Our data suggest that this parietal activation may be better explained by how the brain processes the numeric magnitudes of economic rewards.

Acknowledgments—We thank Kentaro Fujita for providing the Iterated Prisoner’s Dilemma computer program.

REFERENCES


(Rceived 3/18/08; Revision accepted 7/5/08)