Cognitive Constraints on Cooperation

Ellen E. Furlong and John E. Opfer

Research Report

Cognitive Constraints on How Economic Rewards Affect Cooperation

Ellen E. Furlong and John E. Opfer

The Ohio State University

Address correspondence to Ellen Furlong or John Opfer, Department of Psychology, The Ohio State University, 1835 Neil Ave. Columbus, OH 43210, e-mail: furlong.22@osu.edu or opfer.7@osu.edu.

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 its distinctive 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), where 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.

Cooperation—whether sharing the burden of wind resistance in the Tour de France, forming price-fixing cartels in economic markets, or adhering to arms-control agreements in international treaties—often fails to spread among social actors in proportion to such behavior's potential benefits (Olson, 1965). To understand the minds of uncooperative agents, behavioral economists and social psychologists use iterated prisoner's dilemma (IPD) games to examine factors leading to cooperation (Axelrod & Hamilton, 1981; Dawes, 1980; Rachlin, 2003; Rapoport & Chammah, 1965). From this approach, manipulating rewards for defecting versus cooperating in such games can help explain uncooperative behavior in real markets (Fehr & Schmidt, 1999). The validity of this approach, however, relies on the assumption that behavior remains invariant when payoffs are linearly transformed, as when rewards are converted to other units, subdivided into equivalent quantities, or increased over orders of magnitude (Rapoport & Chammah, 1965).

The assumption that cooperative behavior remains constant despite linear transformations in rewards, however, implicitly contradicts findings on how agents represent numeric magnitudes. Just as sensations increase logarithmically with stimulus intensity (Fechner's law), representations of numeric magnitude also increase logarithmically with actual value (Dehaene, 2007). Consequently, in behavioral studies, discrimination of numeric quantities decreases with increasing magnitude (Brannon, 2005; Moyer & Landauer, 1967; Siegler & Opfer, 2003; Starkey & Cooper, 1980). In studies of single-neuron activity, logarithmic scaling of numerosity is also evident in the monkey parietal cortex, where number-tuned neurons lose selectivity with increasing set size (Nieder & Miller, 2004). The human parietal cortex is also activated by tasks requiring numeric comparisons (Piazza, Mechelli, Butterworth, & Price, 2002; Pinel, Dehaene, Riviere, & LeBihan, 2001), as well as by economic games (Bechara, Damasio, Tranel, & Damasio, 2005; Glimcher, 2003; Rilling et al., 2002). One possible reason economic games and number comparisons rely on overlapping brain regions could be that economic decisions (e.g., whether to respond to a $3 vs. a $5 incentive) necessarily involve comparing numeric magnitudes (e.g., 3 and 5), which are represented in accordance with Fechner's law.

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, $F(1, 21) = 6.90$, $p_{rep} = .94$, $\eta^2_p = .25$, and mutual cooperation, $F(1, 21) = 5.40$, $p_{rep} = .91$, $\eta^2_p = .21$, than the dollars condition did. Similarly, the dollars condition elicited greater mutual defection, $F(1, 21) = 9.21$, $p_{rep} = .97$, $\eta^2_p = .31$, and a longer latency to "forgive" the Confederate, or to cooperate after the Confederate's first defection, $F(1, 21) = 4.68$, $p_{rep} = .89$, $\eta^2_p = .18$, 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.

Method

Forty-eight students were randomly assigned to play one of four IPD games; two were identical to the games in Study 1 (“$1,” *n* = 12; “100¢,” *n* = 12) and two were numerically identical to the games in Study 1 but had different units and therefore different economic values (“1¢,” *n* = 12: R = 3¢, S = 0¢, T = 5¢, P = 1¢; “$100,” *n* = 12: R = $300, S = $0; T = $500; P = $100). Subjects in Study 2 played against computers that were programmed with TFT, thus behaving like the student confederates in Study 1; all other procedures were identical to Study 1.

Results and Discussion

Against the hypothesis that the results of Study 1 were due to a preference for dollars or cents, a 2 (units: dollars, cents) × 2 (number: 1, 100) multivariate analysis of variance (MANOVA) on the four indices of player strategy revealed no main effect of unit, $F(4, 41) = .08$, $p_{rep} = .05$, nor did unit interact with number, $F(4, 41) = .23$, $p_{rep} = .16$.

We next examined the effect of numeric value (3 or 300) and economic value ($.03, $3, $300) for each of the four indices. Numerically greater rewards increased individual cooperation (Fig. 3), $F(1, 46) = 6.25$, $p_{rep} = .94$, $\eta^2_p = .12$, which otherwise showed no effect of economic value, $F(2, 45) = 1.45$, $p_{rep} = .69$. For example, changing rewards for mutual cooperation from 3¢ to 300¢ increased individual cooperation rates, but an economically identical change from 3¢ to $3 elicited no gains. The same pattern was evident in rates of mutual cooperation, for which numerically large rewards elicited more mutual cooperation than did numerically small rewards, $F(1, 46) = 11.33$, $p_{rep} = .98$, $\eta^2_p = .20$, with no effect of economic value, $F(2, 45) = 2.76$, $p_{rep} = .84$. Further, numerically large rewards elicited less mutual defection than numerically small ones did, $F(1, 46) = 5.68$, $p_{rep} = .93$, $\eta^2_p = .11$, with mutual defection showing no effect of economic value, $F(2, 45) = 1.66$, $p_{rep} = .71$. Intriguingly, players were very quick to “forgive” defections by the computer ($M = 3.7$ trials, $SD = 3.9$), and neither numeric nor economic value influenced forgiveness, number: $F(1, 46) = 2.58$, $p_{rep} = .79$; value: $F(2, 45) = 1.16$, $p_{rep} = .62$.

Finally, given the history of social motives producing effects on prisoner’s dilemma behavior (e.g., Messick & Brewer, 1983), we last compared the effect of number and type of partner (human, Study 1; computer, Study 2) on behavior. On three of four measures, effect sizes were greater for numeric value than they were for partner type, individual cooperation: $\eta^2_p = .12$ versus .01; mutual cooperation: $\eta^2_p = .20$ versus .001; mutual defection: $\eta^2_p = .11$ versus .05. However, the effect of partner type, $\eta^2_p = .08$, was greater than the effect of number, $\eta^2_p = .05$, for forgiveness, the only variable for which we observed a partner × number interaction, $F(1, 67) = 11.21$, $p_{rep} = .99$, $\eta^2_p = .08$.

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 > 1, and since 300¢/500¢ = \$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) ~ 1 (i.e., temptation to defect and cooperate are nearly equal), whereas ln(3)/ln(5) ~ .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, +1000) or multiplied by (× 0.001 or × 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 no variance in individual cooperation rates, $R^2 = 0$, whereas our logarithmic model, ln(R)/ln(T), accounted for virtually all the variance, R^2 = .97 (Fig. 4). The logarithmic model also accounted for more variance in rates of mutual cooperation, ln(R)/ln(T): R^2 = .71; R/T: R^2 = .07, mutual defection, ln(R)/ln(T): R^2 = .55; R/T: R^2 = 0, and forgiveness, ln(R)/ln(T): R^2 = .42; R/T: R^2 = .02, than did the linear model.

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; Rapoport & Chammah, 1965). Thus, when confronted with larger rewards for unilateral defection ($5) than for mutual cooperation ($3) on a trial in an IPD, players often choose the economically larger reward, suggesting that linear transformations of monetary value (e.g., converting $5 to 500¢) would not change cooperation (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 (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 (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¢ to $1) had no impact on cooperation. Nor were effects of number isolated to social situations: Numeric effects 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," 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 IPD computer program.

REFERENCES

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

Bechara, A., Damasio, H., Tranel, D., & Damasio, A.R. (2005). The Iowa Gambling Task and the somatic marker hypothesis: Some questions and answers. *Trends in Cognitive Science*, *9*, 159–162.

Bernoulli, D. (1954). Exposition of a new theory on the measurement of risk (L. Sommer, Ed. & Trans.). *Econometrica*, *22*, 22–36. (Original work published 1738)

Brannon, E.M. (2005). What animals know about numbers. In J.I.D. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 85–107). New York: Taylor and Francis.

Camerer, C., Loewenstein, G., & Prelic, D. (2005). Neuroeconomics: How neuroscience can inform economics. *Journal of Economic Literature*, *43*, 9–64.

Colby, C.L. (1996). A neurophysiological distinction between attention and intention. In T. Inui & J.L. McClelland (Eds.), *Attention and performance XVI* (pp. 157–177). Cambridge, MA: MIT Press.

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

Dehaene, S., (1997). *The number sense: How the mind creates mathematics.* New York: Oxford University Press.

Dehaene, S. (2007). Symbols and quantities in parietal cortex: Elements of a mathematical theory of number representation and manipulation. In P. Haggard, Y. Rossetti, & M. Kawato (Eds.), *Sensorimotor foundations of higher cognition* (Attention and Performance XXII, pp. 527–574). Cambridge, MA: Harvard University Press.

Fehr, E., & Schmidt, K.M. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics*, *114*, 817–868.

Glimcher, P.W. (2003). *Decisions, uncertainty, and the brain: The science of neuroeconomics*. Cambridge, MA: MIT Press.

Glimcher, P.W., Dorris, M., & Bayer, H. (2005). Physiological utility theory and the neuroeconomics of choice. *Game and Economic Behavior*, *52*, 213–256.

Henik, A., & Tzelgov, J. (1982). Is three greater than five? The relation between physical and semantic size in comparison tasks. *Memory and Cognition*, *10*, 389–395.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*, 263–291.

Messick, D.M., & Brewer, M.B. (1983). Solving social dilemmas. In L. Wheeler & P. Shaver (Eds.), *Review of personality and social psychology* (Vol. 4, pp. 11–44). Beverly Hills, CA: Sage.

Moyer, R.S., & Landauer, T.K. (1967). Time required for judgments of numerical inequality. *Nature*, *215*, 1519–1520.

Nieder, A., & Miller, E.K. (2004). A pariteo-frontal network for visual numerical information in the monkey. *Proceedings of the National Academy of Sciences, U.S.A.*, *10*, 7457–7462.

Olson, M. (1965). *The logic of collective action: Public goods and the theory of groups*. Cambridge, MA: Harvard University Press.

Piazza, M., Mechelli, A., Butterworth, B., & Price, C.J. (2002). The quantifying brain: Functional neuroanatomy of numerosity estimation and counting. *Neuron*, *44*, 547–555.

Pinel, P., Dehaene, S., Riviere, D., & LeBihan, D. (2001). Modulation of parietal activation by semantic distance in a number comparison task. *Neuroimage*, *14*, 1013–1026.

Rachlin, H. (2003) Rational thought and rational behavior: A review of bounded rationality: The adaptive toolbox. *Journal of Experimental Analysis of Behavior*, *79*, 409–412.

Rapoport, A., & Chammah, A.M. (1965). *Prisoner's dilemma: A study in conflict and cooperation*. Ann Arbor, MI: The University of Michigan Press.

Riling, J., Gutman, D.A., Zeh, T.R., Pagnoni, G., Berns, G.S., & Kilts, C.D. (2002). A neural basis for social cooperation. *Neuron*, *35*, 395–405.

Siegler, R.S., & Opfer, J.E. (2003). The development of numerical estimation: Evidence for multiple representations of numerical quantity. *Psychological Science*, *14*, 237–243.

Starkey, P., & Cooper, R.G. (1980). Perception of numbers by human infants. *Science*, *210*, 1033–1035.

Stevens, S.S. (1961). To honor Fechner and repeal his law. *Science*, *13*, 80–86.

TABLE 1

The Predictions Made by the Linear and Logarithmic Models for the Five Matrices Used in Study 3.

Condition/Matrix	Linear Model: R/T	Logarithmic Model: ln(R)/ln(T)
"1": R = 3; S = 0; T = 5; P = 1	0.6	0.68
".001": R = 0.003; S = 0; T = 0.005; P = 0.001	0.6	1.10
".01": R = 0.03; S = 0; T = 0.05; P = 0.01	0.6	1.17
"101": R = 103; S = 100; T = 105; P = 101	0.98	1.0
"1001": R = 1003; S = 1000; T = 1005; P = 1001	0.99	1.0

R = reward for mutual cooperation; T = temptation to defect; S = sucker's reward; P = punishment for mutual defection.

Fig. 1. Typical matrix values in the prisoner's dilemma game. The game is defined by a mathematical relation between payoff values such that the temptation to defect (T) is greater than the reward for mutual cooperation (R), which is greater than the punishment for mutual defection (P), which is in turn greater than the "sucker's reward" (S)—when one has cooperated and one's partner has defected.

Fig. 2. Cooperative behavior (measured by number of trials showing individual and mutual cooperation), competitive behavior (number of trials in which both players

defected), and forgiveness latency (number of trials until a player cooperates again after the opponent defects) in the prisoner's dilemma game when rewards were in dollars versus when rewards were an equivalent monetary value in cents.

Fig. 3. Individual cooperation, mutual cooperation, defection rates and forgiveness latencies when values were numerically low (3¢ and $3 conditions) versus numerically high (300¢ and $300 conditions).

Fig. 4. Variance in cooperation rates accounted for by the logarithmic model, ln(R)/ln(T) (left; R^2 = 0.97) versus by the linear model (R/T) (right; R^2 = 0) in Study 3.