The Effects of Reducing Information on a Modified Prisoner's Dilemma Game

Jay C. Brown (Jaybrown@andrew.cmu.edu)
Carnegie Mellon University
Department of Psychology
Pittsburgh, PA 15213

Marsha C. Lovett (Lovett@andrew.cmu.edu)
Carnegie Mellon University
Department of Psychology
Pittsburgh, PA 15213

Abstract

Participants played a modified prisoner's dilemma game in which competition was created using a single player. The competition was between the player at the moment and the player in the future. The complexity of the game was increased across experiments. The transition from Experiment 1a to 1b saw the removal of information about future consequences and past behavior. Experiment 2 removed information about current outcomes. As the complexity of the game increased (both quantitatively and qualitatively) and therefore the external validity increased, the ability to "solve" the game decreased.

Introduction

Many choices we make are between things that make us feel good at the moment and things that are actually better for us in the long run. Choosing to consume alcohol at a party certainly feels better (at the moment) than not consuming alcohol, but in the long run (hangovers, loss of peer respect, etc.) we are certainly better to refrain from this consumption. When purchasing an automobile, assuming equal price, a sports car is definitely flashier than a mini-van, but the mini-van will probably last longer, be more practical, and cost less in insurance. Purchasing the sports car may make us feel better at the moment, but the total utility (over the life of the vehicles) would unquestionably be higher for the mini-van.

Impulsiveness is defined as the choice for the outcome that feels good at the moment (consuming alcohol; the sports car). Self-control is the choice for the outcome that is actually better in the long-run (refraining from alcohol consumption; the mini-van). Many factors affect our impulsiveness and ability to exhibit self-control. The experiments presented here address several of them.

A goal of this paper is to explore how people learn to choose between impulsivity and self-control. One can view this learning as an adjustment of strategy choices after feedback. Given the task studied in this paper, another relevant perspective views the learning as a growing understanding of cooperation in an iterated prisoners dilemma game.

In a traditional prisoner's dilemma game, two players each choose between two options (often called cooperate and defect) creating four possible outcomes (Rapoport & Chammah, 1965). These outcomes are associated with different rewards, labeled A, B, C and D, that must obey the following rules:

$$B > A > D > C$$
$$2A > B + C > 2D$$

Both players will receive outcome A (moderately good) if both choose to cooperate. Both players will receive outcome D (moderately bad) if both choose to defect. However, if Player 1 chooses to cooperate and Player 2 chooses to defect, then Player 1 will receive outcome C (the worst) while Player 2 will receive outcome B (the best). Defection tends to dominate in both one-shot and iterated playing of this game. However, if the players know each other, and, more importantly, trust each other, then cooperation can arise and persist. Rachlin, Brown and Baker (2001) have shown that Player 1 will cooperate only if he or she believes Player 2 will reciprocate that cooperation.

This result suggests that converting the two-player version of the game to a single-player version[1] would lead to high levels of cooperation. Nevertheless, previous work has shown that individual players chose to "cooperate" with themselves only 54% of the time (Brown & Rachlin, 1999). Why is this percentage so low? Consider the competition engendered by the single-player game: it is between the self at the moment and the self in the future, which is essentially a choice between impulsivity and self-control. The present experiments explore the processes by which individuals choose between these options in the single-player game.

Experiment 1a

The first experiment was designed to be a computer-based replication of previous work (see Brown & Rachlin, 1999 for full details) to obtain greater control

---

[1] A single-player prisoner's dilemma game, where that player makes two or more sequential choices, is identical to a two-player game in which the "other" player uses the tit-for-tat strategy perfectly (see Axelrod, 1987). In both instances, levels of uncertainty exist for the player at the moment of choice as to the future outcomes of the game. Whether that uncertainty arises from a lack of knowledge of another player's future actions or one's own future actions is inconsequential.

of delays and to obtain information on the amount of time participants took to make choices, that is, to increase internal validity.

Method

Participants

Fifty undergraduate students (23 males and 27 females) from the Carnegie Mellon University subject-pool participated in this experiment.

Apparatus and Procedure

All experimental stimuli were presented on an iMac computer with a 14 in. screen (13 in. viewable) using the cT programming language. Participants' choices were made using mouse clicks at the appropriate areas of the screen.

Participants were first asked a series of demographic questions including gender, age, SAT scores, etc. Following this, participants were shown an instruction screen that read as follows:

> You will be playing a game on the board shown on the left [See Figure 1] in which you will be using keys to open doors. These keys and doors will be Red and Green. Red keys open Red doors. Green keys open Green doors. At any given time you will possess a single key. This key can be used to open the appropriately colored door. Whenever you use a key to open a door, you will give up that key. Upon opening a door, you will receive the number of points contained in the box with the door, and a new key. Choices will be made by pressing buttons which will appear on the doors. While you are playing the computer will always show you the last choice you made and the number of points you have earned. You will be given a Red key to begin the game, but after that, the key you have will depend on your actions. Throughout the game, the board will look exactly as it does now. Your goal in playing this game is to earn as many points as you can.

A series of screens followed giving examples of the rules. Post-experimental questioning revealed that all participants understood the rules and procedures of the game upon reading the instructions.

Following the instructions, participants began the actual choice procedure. To avoid procedural errors during the choice procedure, only the currently valid choices (the top doors if the participant possessed a red key, the bottom two if green) were available for choice.

Participants were given a red key to begin the game and were allowed to choose between the top two doors (choose between 15 and 20 points). Regardless of the row in which the participants were currently choosing, choice for the greater points always led to the bottom of the board on the following trial and choice for the smaller points always led to the top of the board on the following trial. The "solution" to this game is such that choice for the smaller amount at the moment always led to more on future trials.

Participants made 100 choices on the screen shown in Figure 1. Participants were able to see the game board at all times, as well as the key they currently possessed, their previous choice, and the total points they had earned. An inter-trial interval (ITI) occurred between each choice (default value 3 s). During the ITI all text was removed from the screen as were the current key and previous choice information. During the ITI the key from the chosen square moved down to the current key position and the points from the chosen square were moved to the points bucket. Following the ITI of the 50^{th} choice, a 4-min filler task (drawing) was employed. Following the filler task, participants made the final 50 choices.

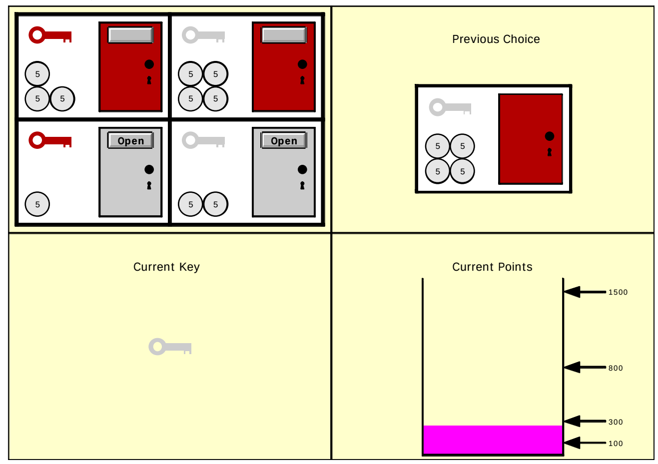

Figure 1: Screenshot of the game. Note: The top two doors are red; the bottom two doors are green; the left compartments contain red keys; the right compartments contain green keys.

Other non-prisoners' dilemma work suggests two features of the game that can impact participants' choices. These are reward amount and ITI. Within the limits of the equations presented earlier, these features can be manipulated. While the ratio of the rewards is the most important factor in choice, absolute levels of the rewards also has a considerable effect (Rachlin, Brown, & Cross, 2000). With regard to the ITI, previous work has shown that engendering commitment to a choice increases cooperation (Rachlin, 1991; Stuart, 1967). One way to accomplish this is by delaying the time until the choice takes place, in other words, increasing ITI.

Experiment 1a manipulated ITI as a between-groups factor to take the values of 3, 6, or 9 s. Additionally, within the 3 s level of ITI, the absolute level of reward was manipulated by a factor of five as a between-groups factor by writing either 5's in the circles of each choice box (as shown in Figure 1) or by writing 1's in the circles.

Results and Discussion

Performance on the iterated prisoner's dilemma game is a function of learning. As such, typical measures of learning, including both latency and performance, were measured. Performance was measured with the percent of trials on which participants cooperated (chose the options that contained fewer points, top left or bottom left box). The first two results address whether cooperation was influenced by ITI or the absolute level of the reward.

An independent-measures ANOVA was performed across the three levels of ITI. Cooperation over all 100 trials did not significantly differ across ITI levels of 3 s (*M* = 86.4%, *sd* = .166), 6 s (*M* = 87.1%, *sd* = .170), and 9 s (*M* = 86.1%, *sd* = .162), $F(2, 47) = .01$, $p > .05$. The absolute level of reward variable was varied only within participants using an ITI of 3 s. An independent-measures *t*-test with these participants revealed that having 5's in the circles (*M* = 86.7%, *sd* = .157) did not engender significantly different cooperation than having 1's (*M* = 85.8%, *sd* = .191), $t(28) = .13$, $p > .05$. Because neither of the manipulated factors had a significant impact on cooperation further testing collapsed across these factors.

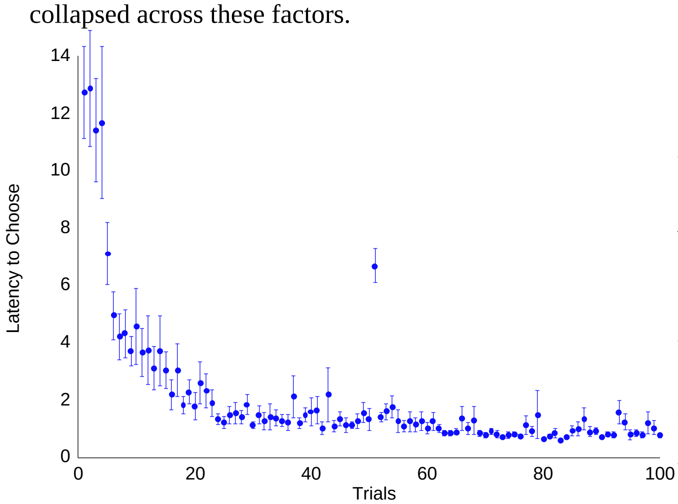

Figure 2: Latency to choose as a function of trial for Experiment 1a. Note: Bars represent ± 1 SE; the single extreme data point at trial 51 occurred immediately following the filler task.

The latencies to choose decreased in an inverse relationship with trials (see Figure 2). The best fitting curve for this data (r^2 = .823) was:

$$Latency(Trials) = 1.7026 / Trials^{.6636}$$

Cooperation by the participants across the 100 trials of Experiment 1a increased logarithmically (see Figure 3). The best fitting curve for this data (r^2 = .634) was:

$$Cooperation(Trials) = 7.203 * \ln(Trials) + 60.25$$

The previously mentioned manual experiment (Brown & Rachlin, 1999) was run for only 40 trials. An independent-measures *t*-test using only the first 40 trials of the present experiment showed that the participants in Experiment 1a (*n* = 50; *M* = 81.6%, *sd* = .170) cooperated significantly more than the participants in the previous experiment (*n* = 20; *M* = 53.75%, *sd* = .097), $t(59) = 8.61$, $p < .001$.

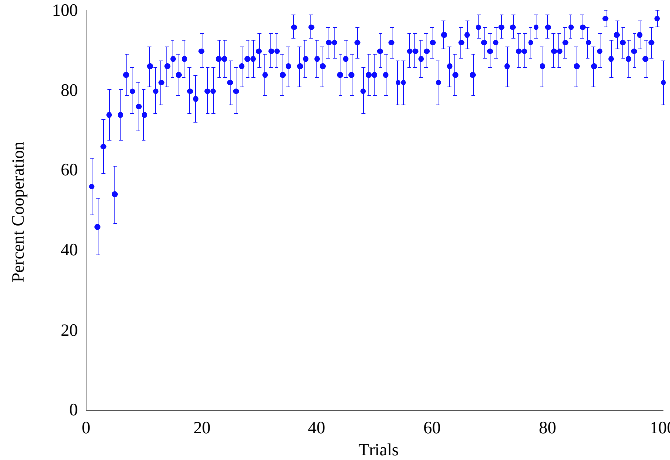

Figure 3: Percent cooperation as a function of trials during Experiment 1a. Note: Bars represent ± 1 SE.

Experiment 1a revealed a ceiling effect on cooperation that had not been seen in previous work. Several possible explanations for this effect exist. First, previous work had been performed using physical apparatus which may be viewed differently from a computer. Second, Experiment 1a was 100 trials long whereas previous work had continued for only 40 trials. However, the ceiling effect was already beginning to show by the 40th trial of Experiment 1a. Third, and perhaps most importantly, the participants in Experiment 1a were extremely analytical (average SAT-M score over 700). Post-experimental questioning revealed that participants treated the experiment as a problem to be solved mathematically.

Experiment 1b was created primarily to investigate the final concern. However, Experiment 1b was also created to increase the external validity of the game.

Experiment 1b

When making choices in life, one rarely knows for certain the consequences of those choices. In fact, it is only through prior experience that we have any idea of future consequences. This experience-based learning was missing from Experiment 1a and was created in Experiment 1b by removing the keys from the boxes.

Method

Participants

Thirty undergraduate students (19 males and 11 females) from the Carnegie Mellon University subject-pool participated in this experiment.

Apparatus and Procedure

Experiment 1b used exactly the same apparatus as Experiment 1a. The ITI in Experiment 1b was 3 s and the circles had 1's written in them. Experiment 1b differed from 1a in that some of the information which had previously been available to the participants was removed from the game. Namely, participants no longer had access to information about their actions on the previous trial and the locations of the keys in the game board (representing future consequences). Additionally, ten of the participants provided a verbal protocol while performing the experiment.

Results and Discussion

An independent-measures *t*-test on the overall percent cooperation showed that participants providing a verbal protocol (M = 83.2%, sd = .172) did not significantly differ from those not providing it (M = 79.9%, sd = .223), $t(28) = -0.41$, $p > .05$. Because of this, further analyses collapsed across this factor.

However, protocol data revealed that participants were still "solving" the game as a math problem. Participants were adding together the outcomes of different combinations of multiple trials, comparing these totals, then merely selecting the combination with the highest total. These participants were again highly analytical (SAT-M average over 700).

In much the same way as Experiment 1a, latency decreased across trials in an inverse relationship. The best fitting curve for this data ($r^2 = .739$) was:

$$\text{Latency}(\text{Trials}) = 8.77/\text{Trials}^{.5468}$$

The cooperation in Experiment 1b increased across trials logarithmically as Experiment 1a. The best fitting curve for this data ($r^2 = .653$) was:

$$\text{Cooperation}(\text{Trials}) = 15.18 * \ln(\text{Trials}) + 25.75$$

A 2 X 10 (Experiment X Ten Trial Blocks) mixed-factorial ANOVA was performed. The overall cooperation in Experiment 1a (M = 86.5%, sd = .163) was not significantly different from Experiment 1b (M = 81%, sd = .205), $F(1, 78) = 1.73$, $p > .05$. The cooperation across the ten ten-trial blocks (M's = 59.9%, 81.3%, 82.4%, 86.4%, 85.9%, 82.9%, 89.4%, 92%, 92.0%, 92.1%; sd's = .263, .242, .240, .229, .240, .216, .204, .185, .174, .166) was significantly different, $F(9, 702) = 53.59$, $p < .001$. A trend-analysis revealed a significant linear component, $F(1, 79) = 76.64$, $p < .001$, suggesting that steady learning occurred across the experiments. Additionally, the interaction of Experiment and trial block was significant, $F(9, 702) = 7.66$, $p < .001$. Planned comparisons showed that the essential differences in Experiments 1a and 1b occurred during the first [$F(1, 78) = 15.86$, $p < .001$] and sixth [$F(1, 78) = 8.19$, $p < .01$] blocks of ten trials in which the participants' in Experiment 1b cooperated less.

Procedurally, Experiment 1a and 1b differed in a quantitative manner. The task was made slightly more difficult by removal of the keys (information about future consequences) and the previous trial reminder (information about past behavior). This quantitative change in task difficulty created only a slight difference in overall cooperation, with its effects felt mainly in the first and sixth block of trials (immediately following the filler task).

Throughout Experiments 1a and 1b, all participants (N = 80) "solved" the game in a sudden fashion using one of several strategies. The first strategy was total cooperation, which was achieved by selecting the top-left door continually. The second strategy, which yielded approximately 66% cooperation, involved choosing the top left door, then the top right door, then the bottom left door, then repeating the sequence. The third common strategy, which created approximately 50% cooperation, involved choosing the top right door, then the bottom left, over and over. Participants using the first strategy (n = 61) tended to do so during the early trials of the experiment (M = 12.3, sd = 15.1). Participants using the second strategy (n = 11) tended to begin it late (M = 40.1, sd = 24.7). Participants implementing the third strategy (n = 8) did so in between (M = 29, sd = 26.9). Once participants moved into one of these strategies, they stuck with them nearly exclusively. Strategy onset was defined using the following two rules. First, the behavior for at least 10 trials following the point of strategy onset must be consistent. Second, the participant must have followed this strategy throughout the game.

A one-way repeated measures ANOVA was performed on the latency to choose on the five trials surrounding the strategy onset (two trials prior to onset, actual onset, and two trials following onset). Eleven participants were removed from this analysis (all using the pure cooperation strategy) because their strategy onset occurred on the 1st or 2nd trial such that they provided no data for trials prior to onset (if included, the effect is more pronounced). These latencies as a function of position, which can be seen in Figure 4, were significantly different, $F(4, 272) = 3.58$, $p < .01$. A planned comparison revealed that the latency on the trial immediately following strategy onset (M = 8.45 s, sd = 15.9) was significantly longer than all other trials averaged together (M = 4.28 s, sd = 7.5), $F(1, 68) = 6.39$, $p < .05$.

Prior to the onset of a steady strategy, verbal protocol data revealed that participants' comments tended to focus on the outcome of the immediate trial. The trial on which strategy onset occurred was treated no differently from previous trials. However, on the trial following strategy onset, the participants would stop and reconsider the entire game, focusing on the game as a "whole", hence, the increased latency on the trial immediately following strategy onset. The results are particularly striking given that earlier trials tend to have a longer latency (see Figure 2 and the downward slope among 4 of the 5 points in Figure 4). This aspect of the data would have the effect of negating the increased latency on the trial following strategy onset.

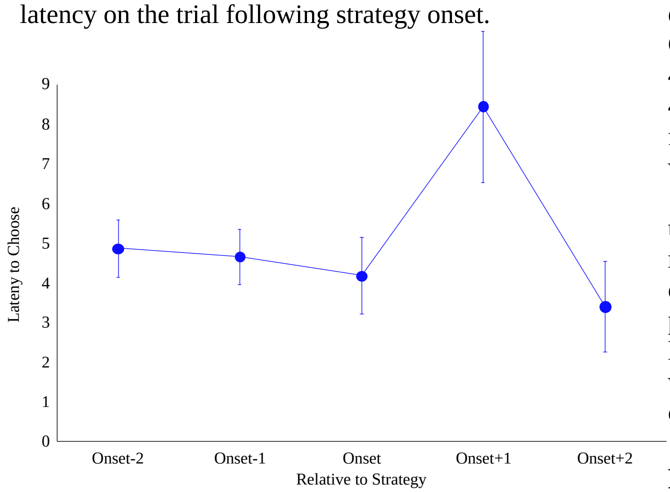

Figure 4: The latency to choose on the trials surrounding strategy onset in Experiment 1a and 1b. Note: Bars represent ± 1 SE.

Experiment 2

Due to the ceiling effects experienced in both Experiments 1a and 1b, participants' treatment of the game as a mathematical problem to be solved was addressed. Real-life choices rarely can be solved mathematically. It is often impossible to verbalize why option A was chosen over option B, it may simply "feel" better. This intuitive nature of decision-making was sought by creating two probability combinations such that reinforcement was unknown in advance rather than deterministic as it had been in Experiment 1.

Method

Participants

Forty undergraduate students (27 males and 13 females) from the Carnegie Mellon University subject-pool participated in this experiment.

Apparatus and Procedure

Experiment 2 used the same apparatus as Experiments 1a and 1b. The ITI was 3 s and the circles had 1's written in them. It will be remembered that in Experiment 1 (see Figure 1) the game board contained 3, 4, 2, and 1 circles in the different compartments. In Experiment 2 all compartments contained 5 circles. However, in Experiment 2, upon making a choice, the receiving of points was probabilistic. If a participant chose any given door, they may or may not receive the 5 points. The probabilities used were manipulated. Participants in Group 1 (*n* = 20) received the 5 points 60%, 80%, 40%, or 20% of the time (starting in the top left and going clockwise). This created expected values for these choices that were 3, 4, 2, and 1 (i.e. 5 points received 60% of the time has an expected value of 3), exactly as they were in Experiment 1. Participants in Group 2 (*n* = 20) received the 5 points 80%, 100%, 40%, and 20% of the time creating expected values of 4, 5, 2, and 1. Due to an apparent lack of asymptote following 100 trials, 12 of the participants in Group 2 were given 100 additional trials for a total of 200.

The previously listed probabilities were generated using a true probability mechanism, However, several rules were employed which ensured both that the obtained probabilities approximated the programmed probabilities at both the local and global level and that long strings of wins or losses (such that the string would be expected to occur less than 5% of the time by chance) were avoided.

Results and Discussion

A 2 X 10 (Group X Block of ten trials) mixed-factorial ANOVA was performed on the cooperation during the first 100 trials. Group 1 (*M* = 53.5%, *sd* = .143) did not cooperate significantly less overall than Group 2 (*M* = 58%, *sd* = .106), $F(1, 38) = 1.28$, $p > .05$. The cooperation across the ten ten-trial blocks did not significantly differ, $F(9, 342) = .36$, $p > .05$. The interaction of Group and block was not significant, $F(9, 342) = 1.2$, $p > .05$.

In a similar manner seen in the differences between Experiments 1a and 1b, the slight difference in results for Groups 1 and 2 seemed to reflect a quantitative change in task difficulty. Group 2's more divergent probabilities, and hence expected values, created more cooperation.

A single-sample *t*-test on the overall cooperation for Group 1 showed that it did not significantly differ from random responding (50%), $t(19) = 1.08$, $p > .05$. A second single-sample *t*-test on cooperation during the final ten-trials (when responding is most stable, *M* = 55.5%, *sd* = .233) revealed the same lack of difference, $t(19) = 1.06$, $p > .05$.

Perhaps the probabilities of reinforcement used with Group 1 were too subtle even though the expected values were identical to the points used in Experiment 1. Inspection of the data shows virtually no change in cooperation from the beginning to the end of the experiment. The effects of continuing the experiment beyond 100 trials remain to be seen.

A single-sample *t*-test on the cooperation for Group 2 during the first 100 trials (*n* = 20) showed that it was significantly higher than chance, $t(19) = 3.35$, $p < .01$. However, cooperation during the 10^{th} block of ten trials

(M = 58.5%, sd = .223) was not significantly higher than expected by chance, $t(19) = 1.7$, $p > .05$.

A single sample t-test on cooperation over all 200 trials (M = 63.4%) for those that received them (n = 12) showed that it was significantly higher than chance, $t(11) = 3.33$, $p < .01$. Likewise, cooperation during the final ten-trial block (M = 73.3%, sd = .227) was significantly higher than chance, $t(11) = 3.56$, $p < .01$.

The cooperation of Group 2 continued a slow growth throughout the first 100 trials of the experiment. This growth also continued through the second 100 trials. One wonders where this group's cooperation would asymptote.

The latency data of Experiment 2 mirrored the inverse relationship seen in both Experiments 1a and 1b. Latency decreased across trials in an inverse relationship. The best fitting curve for this data, which created an r^2 of .896, was:

$$\text{Latency(Trials)} = 5.52 / \text{Trials}^{.4186}$$

An examination of the relationship of cooperation across trials for the participants in Group 1 showed that no function could account for more than 2% of the variance. The relationship between cooperation and trials was slightly better for Group 2. A logarithmic function captured the greatest amount of variance in this relationship creating an r^2 of .151:

$$\text{Cooperation(Trials)} = 5.11 * \ln(\text{Trials}) + 41.52$$

A similar analysis of strategy, as was done in Experiment 1, was attempted for Experiment 2. Even when using a lenient definition of strategy onset, only 16 of the 40 participants were classified as having employed a steady strategy. A one-way repeated measures ANOVA on the latency to choose on the trials surrounding strategy onset revealed no differences, $F(4, 60) = .89$, $p > .05$.

General Discussion

When the internal validity of the game was increased (prior experimentation to the present experiments), participants' ability to cooperate with themselves increased substantially (though population differences may have had a large impact on this change). As the complexity, external validity, and face validity of the game increased within these experiments (both quantitatively and qualitatively), participants' ability to cooperate with themselves decreased.

Participants' behavior in Experiments 1a and 1b was characterized by a sudden change: at one moment focusing on the current trial, at the next moment focusing on the whole game. This "insight learning" showed itself in the strategy results. There was some moment when each participant stopped responding semi-randomly and began responding according to a strategy. In Experiment 2 behavioral change occurred gradually, though slightly faster for Group 2. This "trial and error" learning was shown in two ways. First, by the absence of strategy results. Second, by the participants' use of a win-stay/lose-shift approach.

The procedural differences between Experiments 1a and 1b created a quantitative change in behavior, 1b was slightly more difficult than 1a. Participants in both experiments "solved" the problem in a moment of insight. The procedural differences between Group 1 and Group 2 in Experiment 2 created quantitative differences in behavior, with the more divergent expected values in Group 2 creating more cooperation. Both groups appeared to use a form of trial and error learning. However, the change from Experiment 1 to 2 was qualitative. Participants solved these experiments using completely different approaches.

The insight shown in Experiment 1 occurred only after the participants had started using a strategy. On the trial following strategy implementation, the participants stopped and viewed the entire experiment, adding together the points from various combinations of moves. Because no verbal protocols were used in Experiment 2, it is impossible to know for sure how these participants viewed the problem. Perhaps the participants viewed the game as they would a real life problem. Picking one option "feels" better than picking another. Perhaps a simpler explanation based on probability matching may be a better explanation.

References

Axelrod, R. (1987) The Evolution of Strategies in the Iterated Prisoner's Dilemma. In Lawrence Davis (ed.), *Genetic Algorithms and Simulated Annealing*. London: Pitman.

Brown, J. C. & Rachlin, H. (1999). Self-control and social cooperation. *Behavioural Processes*, *47*, 65-72.

Rachlin, H. (1991). *Introduction to Modern Behaviorism* (3rd ed.). New York: W. H. Freeman and Co.

Rachlin, H., Brown, J., & Baker, F. (2001). Reinforcement and punishment in the prisoner's dilemma game. In D. L. Medin (ed.), *The Psychology of Learning and Motivation-Advances in Research and Theory: Vol. 40*. New York: Academic Press.

Rachlin, H., Brown, J., & Cross, D. (2000). Discounting in judgments of delay and probability. *Journal of Behavioral Decision Making*, *13*, 145-159.

Rapoport, A. & Chammah, A. M. (1965). *Prisoner's Dilemma-A Study in Conflict and Cooperation*. Ann Arbor, MI: The University of Michigan Press.

Stuart, R. B. (1967). Behavioral control over eating. *Behavior Research and Therapy*, *5*, 357-365.