Forgetting constrains the emergence of cooperative decision strategies

Jeffrey R. Stevens[1], Jenny Volstorf[1], Lael J. Schooler[1], Jörg Rieskamp[1,2]

[1]Center for Adaptive Behavior and Cognition, Max Planck Institute for Human Development, Berlin, Germany; [2]Department of Psychology, University of Basel, Basel, Switzerland.

Running title: Forgetting constrains cooperative strategies

Correspondence:
Dr. Jeffrey R. Stevens
Center for Adaptive Behavior and Cognition
Max Planck Institute for Human Development
Lentzeallee 94
14195 Berlin, Germany
jstevens@mpib-berlin.mpg.de

Abstract

Theoretical studies of cooperative behavior have focused on decision strategies that depend on a partner's last choices. The findings from this work assume that players accurately remember past actions. The kind of memory that these strategies employ, however, does not reflect what we know about memory. Here, we show that human memory may not meet the requirements needed to use these strategies. When asked to recall the previous behavior of simulated partners in a cooperative memory task, participants performed poorly, making errors in 10-24% of the trials. Participants made more errors when required to track more partners. We conducted agent-based simulations to evaluate how well cooperative strategies cope with error. These simulations suggest that, even with few errors, cooperation could not be maintained at the error rates demonstrated by our participants. Our results indicate that the strategies typically used in the study of cooperation likely do not reflect the underlying cognitive capacities used by humans and other animals in social interactions. By including unrealistic assumptions about cognition, theoretical models may have overestimated the robustness of the existing cooperative strategies. To remedy this, future models should incorporate what we know about cognition.

Keywords: agent-based simulation, cooperation, decision strategy, forgetting, memory, prisoner's dilemma, tit-for-tat

Introduction

Theoretical analyses have demonstrated that cooperation can evolve in situations in which individuals interact repeatedly and their behavior depends on other's and/or their own past behavior (Axelrod & Hamilton, 1981; Nowak, 2006). For instance, the celebrated decision strategy tit-for-tat (TFT) cooperates on the first move with a partner and then copies the partner's single last choice (cooperate or defect) on all subsequent interactions (Axelrod & Hamilton, 1981; Rapoport & Chammah, 1965). This and similar strategies, such as generous TFT, contrite TFT, tit-for-two-tats, and win/stay-lose/shift (Boyd, 1989; Kraines & Kraines, 1989; Nowak & Sigmund, 1992), have dominated theoretical studies of cooperation for the last 30 years. Despite their dominance in the theoretical work, the assumptions about the underlying cognition required to implement these strategies have not been adequately tested. Thus, there is a critical gap between the theoretical work on which decision strategies can maintain cooperation and the empirical work on what strategies individuals actually use. To bridge this gap, we must test whether the cognitive mechanisms required to implement strategies are psychologically plausible (Hammerstein, 2003; Stephens, McLinn, & Stevens, 2002; Stevens, Cushman, & Hauser, 2005; Stevens & Hauser, 2004; Todd & Gigerenzer, 2007). Here, we investigate one of these cognitive mechanisms: memory for past actions. We ask whether existing strategies make reasonable assumptions about memory or whether problems associated with forgetting could constrain the emergence of these cooperative strategies.

Memory represents a primary cognitive capacity needed for strategies in social interactions that depend on past behavior. The strategies tested in the literature for social interaction make different memory requests. The so-called memory-1 strategies require that players accurately remember the single last choice from each partner. Memory-2 strategies require accurate memory for the last two choices. Humans and other animals, however, sometimes forget. If an individual cannot remember the past action of an interaction partner, then he or she cannot apply a strategy that relies on this knowledge.

In contrast to the existing cooperative strategies, our memory does not work like computer memory, filing away pieces of information for flawless retrieval later. Instead, our memory functions more like how a search engine retrieves information from the internet, with memory records associated to retrieval cues (Anderson et al., 2004; Estes, 1955), much like how websites are indexed by keywords. This associative nature of memory leads to problems of interference, in which cues become associated with many memories, hindering the retrieval of the information sought. Our memory suffers from both proactive interference, in which old memories disrupt the retrieval of new information, and retroactive interference, in which new memories disrupt retrieval of old information.

Despite its central importance, the role of memory in cooperation has received little attention in the existing literature. In one of the few studies to explore memory and cooperation, Milinski and Wedekind (1998) examined the effects of working memory on cooperation by giving half of their participants a working memory task between interactions. They found that without the memory task, participants seemed to use a more complicated memory-2 strategy, whereas with the memory interference, they used a simpler memory-1, TFT-like strategy. Winkler and colleagues (2008) introduced multiple partners to track, as well as varied the interaction pattern between repeatedly interacting the same partner

before switching to a new one or randomly interacting with partners. When randomly interacting with partners, participants with better recall of biographical information about their partners received higher payoffs in the cooperative games—better memory abilities at the individual level resulted in higher payoffs. These studies either measured or manipulated memory performance for information outside of the cooperative situation. Here, we test the role of memory for partners' previous actions on cooperation.

Given the nature of memory, we ask whether existing decision strategies that promote cooperation (such as TFT and its variants) are cognitively feasible. We explore whether humans have the memory capacity required to implement these strategies. Thus, we are asking whether individuals *can* use strategies like TFT, not whether they *do* use these strategies. Do they have the requisite cognitive capacity? To address this capacity question, we designed a memory experiment that tested the role of memory interference on the ability to recall past actions. Though this study does not mimic real world cooperative situations, it was not meant to. Our experimental design replicates the conditions under which memory-1 and -2 strategies should work in order to test the underlying cognitive assumptions of these strategies.

We conducted an experiment with human participants, in which a series of simulated partners chose to cooperate or defect. We measured participants' memory accuracy in recalling each partner's last action. To test the effects of memory interference on cooperation, we implemented two experimental manipulations. First, we varied the number of simulated partners, which is critical when interactions between different partners are interleaved (e.g., partner A, partner B, partner C, partner A, etc.). In this case, an individual may forget a specific partner's previous behavior due to the intervening interactions interfering with the retrieval of the memory; more partners result in more retroactive interference. Second, we varied the number of interactions with each partner because more previous interactions might interfere with the ability to recall only the single last interaction (proactive interference). From these manipulations, we can estimate how memory errors respond to increases in proactive and retroactive interference.

Estimates of memory accuracy alone, however, do not demonstrate the complete role of memory in cooperation. We must also test how well specific decision strategies cope with error caused by misremembering a partner's last actions. For instance, TFT's performance decreases when errors exist because mistakenly defecting results in the lower payoffs of mutual defection (Molander, 1985). A more forgiving form of TFT called contrite TFT (CTFT; Boyd, 1989) outperforms TFT when individuals make errors. Although a few strategies have been tested over a few error rates (e.g., Rieskamp & Todd, 2006; Stephens, Nishimura, & Toyer, 1995; Wu & Axelrod, 1995), to our knowledge there exists no comprehensive treatment of error on the memory-1 and -2 strategies. We used agent-based simulations to systematically analyze the success of several strategies proposed in the literature across a broader range of error rates. With these and the human memory results in hand, we can determine whether currently proposed decision strategies provide adequate models of cooperative behavior.

Cooperative memory experiment

Methods

We recruited 216 participants (age: mean±standard deviation = 25.4 ± 3.2 years, range = 18-36 years) drawn from Berlin universities via the Max Planck Institute for Human Development participant pool. We prepared all materials in German and programmed the experiment in E-prime experimental software (Schneider, Eschmann, & Zuccolotto, 2002). The program began by asking participants to provide demographic data (sex, age, educational level, occupation, college major).

Before beginning the experiment, participants received a paper copy of instructions (Appendix A) describing the goal of the task: recall the last action (cooperate or not cooperate) for each simulated partner. Participants returned the instructions to the experimenter before continuing to avoid giving them a means to record information during the task. A practice phase familiarized participants with the experiment. The practice phase was identical to the actual experimental session, except (1) it used fewer trials in a fixed order for all participants (three partners with four interactions each and six partners with three interactions each), (2) it included only female partners (the experimental phases included only male partners), and (3) the money earned did not accumulate for the final payment. At the end of the practice session, participants received feedback concerning their success ("You have accomplished the practice session with x out of 21 correct answers.").

Following the practice phase, participants experienced one of the nine experimental conditions (24 participants—12 males and 12 females—in each condition) that differed in the number of simulated partners per group (5, 10, or 15 partners) and the number of interactions with each partner (5, 10, or 15 interactions). To keep the number of trials as similar as possible for each participant, we replicated some of the conditions several times until the participants experienced between 150-225 trials. Thus, some conditions had only one replicate, whereas others had up to six replicates (Table 1).

Each replicate consisted of a series of rounds, each with a different set of partner names and images. Participants met with each partner once in a randomized order per round. In the first round, we presented individually for each partner a photograph, a name, and an action: for instance, "Klaus cooperates" (Figure 1). All partners were male, and we randomized partner names and photos across participants. Participants viewed each partner's information for 5 s before advancing to the next partner (1 s in between partners). For every trial in the experimental phase, we randomly assigned the partner's action as cooperate or defect, so participants could not associate a pattern of action with each partner and had to track the exact behavior of each partner in the previous round.

After viewing all members of one group, participants began the retrieval rounds, with a randomized order of partners in each round. We presented the image of the partners, along with the question "What did [name] do last time?". The participant had 10 s to answer by pressing "k" or "n" ("*kooperiert*" ["cooperate"] or "*nicht kooperiert*" ["did not cooperate"]) on the keyboard. If they responded within 10 s, they received a feedback screen for 3 s stating whether they were correct, the amount of money they received for that trial (only if they were correct), and an updated total amount received so far in the experiment. If they failed to respond in time, the participant did not receive feedback, only a reminder to respond more quickly next time. After the feedback screen, participants viewed the new action of the

current partner for 5 s before advancing to the next partner. In between rounds, participants could pause the program and start a new round at their discretion. Afterwards, participants completed a questionnaire asking what kinds of strategies they used to solve the memory task, as well as how often they guessed and how often they thought the partners cooperated. Participants received 0.05 euros for each correct answer and 5 euros for showing up, earning an average of 11.05 euros (approximately 14 US dollars) per person (range = 8.25-14.60 euros). We analyzed the data using R statistical software version 2.15.1 (R Development Core Team, 2010) and the epicalc (Chongsuvivatwong, 2010), Hmisc (Harrell, 2010), and lattice (Sarkar, 2008) packages. The original document for this paper used Sweave (Leisch, 2002) to embed the R code into the document, thus ensuring reproducible research (Leeuw, 2001). Data and R code are available in the Supplementary Materials.

For the photographs of partners, we used images from Ebner (2008) downloaded from the Center for Vital Longevity: http://vitallongevity.utdallas.edu/stimuli/facedb/categories/neutralized-faces-by-natalie-ebner.html. We used 9 images of females for the practice phase and 31 images of males for the experimental phase. The depicted persons ranged between 18 and 32 years old, with the same background and color of clothing (Ebner, 2008). For partner names, we used 40 of the most common male German names from 1958-2000, retrieved from http://www.gfds.de/vornamen/beliebteste-vornamen/ (Figure 1).

Results

As shown in Figure 2, participants made more errors as group size increased. With a group size of 5 partners, participants made errors in a mean (±95% confidence interval) of 9.5±2.3% of trials, whereas with 10 and 15 partners, they made errors in 22.5±2.5% and 24.0±2.5% of trials respectively. Participants performed fairly accurately at the smallest group size, but once required to track 10 or more partners, memory errors increased dramatically. In fact, the error rates in the 10- and 15-partner conditions suggest that participants were guessing in half of the trials. Thus, retroactive interference from tracking multiple partners sharply increased memory errors in this task.

To further explore this memory interference, we examined error as a function of the number of intervening interactions. Between consecutive presentations of the same partner, there were other intervening partners. Because we randomized the order of presentation of partners within a round structure, we had variation in the number of intervening interactions and could test whether more intervening events resulted in worse memory performance. When consecutive interactions with the same partner occurred with no intervening interactions, participants performed well, with a mean error rate below 10% (Figure 3). With even one intervening interaction, however, error rates doubled. With more intervening events, errors continued to increase but at different levels for 5 partners compared to 10 and 15 partners.

With these data, we could estimate a function describing how forgetting increased with the number of intervening interactions. When combining the participants experiencing 10 and 15 partners, these data were well described by the power function $p = 1 - 92(1 + n)^{-0.08}$ ($R^2 = 0.90$), where p represents the probability of an error and n represents the number of intervening interactions. A similar analyis on the 5-partner data yield the power function $p = 1 - 96(1 + n)^{-0.04}$ ($R^2 = 0.90$). We used a modified version of Wickelgren's (1974)

function because it predicts memory data well (Wixted & Carpenter, 2007).

We also examined whether experiencing 5, 10, or 15 interactions with each partner influenced error rates. Surprisingly, the number of interactions did not influence performance (Figure 2). An examination of the trend in error rates across the course of the experimental session suggests a general learning effect. Participant errors increased in early rounds, indicating that more interactions caused more mistakes (Figure 4). Yet, in later rounds, performance almost returned to first-round levels, perhaps due to the participants' developing particular mnemonic strategies. In a questionnaire after the experiment (Appendix B), we asked participants to describe any strategies that they used during the cooperative memory task. A common strategy was to memorize either the cooperators or defectors and then infer the other. Also, participants frequently tried to focus on either positive (for cooperate) or negative (for defect) features of the faces or applied additional letters to the names (e.g, when Tim cooperates, remember Timk or Timko). Some elaborate strategies generated stories (e.g., "I eventually imagined that all the cooperating partners were members of my 'gang' and tried to talk myself into disliking the 'traitors' "). It appears as though participants used specific strategies to help in recall, which may account for the decrease in error rates over trials.

Males and females did not differ in their error rates (males: 19.2±0.6%; females: 18.8±0.6%), and participants experienced similar error rates for cooperation and defection actions (cooperation: 19.2±0.6%; defection: 18.8±0.6%), suggesting no preferential memory for defectors or "cheaters" in this context (Cosmides & Tooby, 1989; Mealey, Daood, & Krage, 1996).

Because both the images and names used as stimuli in this experiment varied in terms of attractiveness (Ebner, 2008; Rudolph, Bohm, & Lummer, 2007), we examined the mean memory performance aggregated over all participants for the images and names that were rated for attractiveness. Attractiveness, however, did not correlate with memory performance for the images (N = 40, r =-0.21, p =0.19) or names (N = 19, r =-0.10, p =0.69).

Simulation analysis

Methods

We conducted a set of agent-based simulations in Pascal (code is available in the Supplementary Materials) in which each agent interacted in a series of repeated prisoner's dilemma games (Table 2). In the simulations, agents used one of nine strategies in the interactions (Table 3): always cooperate (ALLC), always defect (ALLD), contrite TFT (CTFT), generous TFT (GTFT), grim trigger (GRIM), random (RAND), tit-for-tat (TFT), tit-for-two-tats (TF2T), and win-stay/lose-shift (WSLS), also known as Pavlov. The population consisted of 11 agents of each strategy type, resulting in 99 total agents. Based on one of the conditions from the experiment, agents interacted with 10 randomly chosen partners for 10 interaction rounds. After completing all interactions, we summed the payoffs over all interactions for each agent in the population. To generate a new population, we ranked all agents by their total fitness and accumulated the total population fitness, starting at the lowest-ranked agent. We then randomly chose (with equal probability) one number from 0 to the accumulated population fitness. The strategy of the agent associated with that

randomly chosen number was added to the next generation. We repeated this procedure (with replacement) until we populated the next generation with 99 agents. In 2% of the reproductive events, we randomly mutated the chosen strategy to one of the eight other strategies. We continued to produce new generations until all agents in a population played a single strategy. Simulations stopped when the entire population consisted of one strategy.

We introduced error into the simulation by varying the probability of an agent "misremembering"—that is, remembering that the partner chose the opposite of what it actually chose—in six of the strategies: CTFT, GRIM, GTFT, TFT, TF2T, and WSLS. For strategies using multiple previous actions from the partner (CTFT, TF2T), each memory had an independent probability of error. No memory was necessary for ALLC, ALLD, and RAND, and we assumed perfect memory for the agent's own action in CTFT and WSLS. We varied the error rate from 0-50% in 1% increments and conducted 1,000 simulations at each of the 51 increments. We report the proportion of the 1,000 replications in which each strategy dominated the population (i.e., the remaining strategy in the final generation).

Results

In the cooperative memory task, even when explicitly rewarded for recalling the last action of their partners, participants made mistakes in 10-24% of trials. Though these error rates seem quite high given that chance performance in this task is 50%, we need a criterion for determining whether decision strategies can maintain cooperation at the error rates demonstrated by our participants. To determine whether the existing decision strategies can cope with this level of error, we assessed how well these strategies performed when making mistakes in an agent-based simulation. Figure 5 shows for each error rate 1) the performance for each strategy (mean proportion of simulations in which each strategy outcompeted all other strategies) and 2) the proportion of interactions in the last generation in which the agents cooperated. At low error rates, GRIM—a strategy that begins by cooperating, then permanently switches to defection following the partner's first defection—outperformed all other strategies. Though at odds with Axelrod and Hamilton's (1981) original results, this finding replicated results from Linster (1992) in which GRIM dominated the populations in the absence of errors. Additionally, ALLD, WSLS, TFT, and CTFT won a small percentage of the simulations. As error rates increased, ALLD and GRIM outcompeted TFT and the other cooperative strategies. The frequency of cooperative acts employed by all agents in the population decreased dramatically as errors became more prevalent. This decrease in cooperation reflected how the various strategies such as GRIM switched from cooperating to defecting when memory errors increased. Thus, cooperation could not be sustained, even at low levels of error.

To further assess the role of error on cooperation, we embedded the forgetting functions from our experimental data into the agents in our simulation. Instead of using a fixed error rate as in the previous simulation, we conducted a simulation in which the error rate depended on the number of intervening interactions, and we drew that error rate from the fitted forgetting function from the memory experiment. All other aspects of the simulation were the same as above, and we conducted 1,000 replications of this simulation.

Using this forgetting function to assign memory error as a function of number of intervening events yielded results similar to the fixed-rate analysis. ALLD won around

83.0% of the simulations while GRIM won 17.0%, and only 6.2% of interactions involved cooperation. Even when using a lower-error forgetting function based on the 5-partner condition of the experiment, only strategies ALLD and GRIM performed well (winning 74.5% and 24.5% of the simulations, respectively), and we observed cooperation rates of 12.8%. The cooperative strategies that depend on memory of partners' last action failed when confronted with a realistic, forgetful memory.

Game theoretical analysis

To verify our agent-based simulation results, we also used analytical methods to assess the role of error on cooperation by applying evolutionary game theory (Maynard Smith, 1982). Evolutionary game theoretical analyses seek an evolutionarily stable strategy (ESS), that is, a strategy that, when adopted by all members of a population, cannot be outperformed (or invaded) by any alternative strategy. If a strategy A playing against itself has a higher payoff than any alternative strategy has against A (payoff(A, A) > payoff(alt, A)), that strategy A is an ESS. If the payoffs are the same, then A must have a higher payoff against the alternative strategy than the alternative strategy has against itself (payoff(A, alt) > payoff(alt, alt)) to be an ESS. Because we are interested in how error influences the payoffs of many strategies, we used Stephens and colleagues' (1995) technique to calculate ESSes with error. This technique, however, only applies to strategies that use information from the single last interaction. Including earlier interactions greatly complicates the analysis, so we limited this analysis to the seven strategies that use only the last interaction: ALLC, ALLD, GRIM, GTFT, RAND, TFT, WSLS (Table 3). We used the standard prisoner's dilemma matrix (Table 1) and set the probability of future interaction to $\alpha = 0.9$ to approximate the 10 interactions used in our experiment. To estimate the payoffs for the remaining strategies (CTFT and TF2T), we used an agent-based simulation with two agents (one was either CTFT or TF2T and the other was one of the nine strategies) playing 10 interactions for 10,000 replicates. We calculated or simulated the payoffs to each strategy against each other strategy with error rates ranging from 0-50% in 1% increments.

We corroborated the simulation finding with a game theoretical analysis. Figure 6 illustrates the game-theoretical payoffs of all strategies categorized by the strategy against which the others play (the "population" strategy). When the payoffs of a strategy playing against itself exceed the payoffs of all other strategies against it, the strategy is an ESS for these error rates. ALLD was an ESS over the entire range of error rates. GRIM was an ESS at error rates between 12-18%, validating its performance in the evolutionary simulation around that error rate (Figure 5). CTFT was an ESS at error rates between 0-17%, although these results are simulated and must be viewed with caution. Otherwise, none of the other strategies was evolutionarily stable for this range of parameters.

Discussion

The goal of this study was to test the psychological plausibility of the memory assumption implicitly embedded in models of decision strategies for repeated social interactions. These strategies assume that behavior in a social interaction depends on the precise recall of a partner's past actions. We show that human participants have great difficulty accurately recalling the previous actions of simulated partners. Interference associated with tracking

the behavior of partners degrades memory performance, and having more partners results in worse performance. To assess whether the decision strategies proposed in the literature can sustain cooperation in the face of error, we conducted simulations of a repeated prisoner's dilemma game in which the agents sometimes forgot their partner's past actions. When mapping the experimental results onto the simulation results, we see that, in our simulation scenario, cooperation is not maintained because few cooperative strategies perform well at the error rates shown by the experimental data. Instead, defection dominates with these estimates of error. These results held even when we used estimates of the best memory performance observed in our memory experiments. Of course, the results of the simulations are dependent on the strategies included and the parameters used. Nevertheless, these findings support the notion that a complete understanding of cooperation requires investigating the underlying cognition needed to implement those strategies (Hammerstein, 2003; Stephens et al., 2002; Stevens et al., 2005; Stevens & Hauser, 2004; Furlong & Opfer, 2009).

One limitation of our experiment is the artificial nature of the task, a limitation shared by most other cooperation experiments in psychology and economics. More realistic social interactions might trigger more effective memory performance, so we should pay careful attention to how cooperation evolves with error rates lower than what we observed. Though aspects of the task may be artificial, in some ways, our memory task actually underestimates error. For instance, we use rather small group sizes, ranging from 5-15 individuals. Estimates from Christmas card lists in England suggest average social network sizes around 125 individuals (Hill & Dunbar, 2003). Tracking the behavior of this many individuals is quite daunting and likely would greatly increase the error rate. Additionally, our study minimizes the influence of events outside of the cooperative interactions on memory accuracy. In more realistic settings, many more aspects of real life may interfere with accurate memory. We asked participants to recall behavior after rather short delays and with only a few intervening events. In our day-to-day lives, we constantly encode memories that may interfere with our ability to recall, with retention intervals extending into days, weeks, months, or even years between interactions. More realistic situations with larger numbers of social partners and longer time delays between interactions could actually make memory worse than that observed in our study. Thus, our task may be too difficult in some ways and too easy in others, but in either case, the strategies in question need to track behavior with an exquisite memory.

Most empirical studies of the prisoner's dilemma involve repeated interactions with the same opponent. We created a more realistic situation by including multiple partners and interleaving interactions among partners (Winkler et al., 2008). A further improvement might be to offer a skewed interaction pattern. Rather than meeting all partners the same number of times, participants could have interacted more frequently with some partners than others, a pattern we observe in natural social encounters (Pachur, Schooler, & Stevens, in press). These patterns of contact have interesting implications for cooperation because the frequency of contact influences the expected time between contacts. Thus, retention intervals vary for tracking the previous behavior of more versus less frequently contacted social partners.

Finally, in our task, we attempted to make the cooperation and defection events equally salient, but real cooperative interactions are much more heterogeneous: opening a door for someone will not be remembered in the same way as cheating on a spouse. The

salience or magnitude of costs or benefits of the cooperative or defection event likely contributes to the retention of the memory (Mealey et al., 1996; Rankin & Eggimann, 2009). Yet, our analysis with lower error rates (forgetting function based on the 5-partner condition) still showed minimal cooperation rates, indicating that better memory performance is not enough to sustain cooperation—near perfect memory is required. More importantly, we designed a task that is ecological valid for TFT and the other decision strategies under investigation. These strategies do not invoke emotional salience or differential encoding of behavior depending on the magnitude of costs or benefits. They all simply store a binary value (cooperate or defect) for each partner. Adding salience and magnitude effects means developing and testing new strategies, a path we fully endorse.

How might we circumvent the problem of memory in cooperation? Or, put another way, why do we see cooperation in iterated prisoner's dilemma situations? There are at least two possibilities. The first is methodological. Many studies of the prisoner's dilemma have participants play against a single partner repeatedly. This may facilitate cooperation both because it provides much experience with a partner and because it limits the memory load associated with the more realistic scenario of tracking multiple partners. The second reason why we may see cooperation in these tasks is that people are using different strategies than those currently proposed in the literature. One possibility is a kind of meta-strategy in which people use TFT when they can remember past interactions and use another strategy when they cannot remember. Though this meta-strategy has not been investigated theoretically, people could use something like this to reciprocate. Alternatively, people may be using a longer-term reciprocal strategy. Instead of relying solely on the most recent behavior when cooperating, they may build a reputation for partners, accounting for experience over several interactions (Roberts, 2008). People may implement reciprocal strategies that classify partners into types instead of track all individual choices. Though we focused here exclusively on direct experience with partners, people also likely use indirect experience by observing third-party interactions to build an image score for potential partners (Roberts, 2008; Rankin & Eggimann, 2009). Thus, instead of tracking individual interactions, people may encode more general summaries of behavior, drawn from both personal experience and observing other interactions.

Rather than test how people actually make cooperative decisions, our intention here was to test whether the current decision strategies provide a suitable framework for exploring cooperation. We suggest that, though these models have proven valuable in investigating cooperation for the last 30 years, they do not accurately reflect underlying cognition. Humans certainly use reciprocal strategies when cooperating, but they likely do not use strategies like TFT and its relatives. Our results suggest that they simply cannot use these strategies because the memory load is too great. To examine the types of reciprocal strategies that humans and other animals use, we must embed what we know about memory into new realistic cooperative strategies. Building psychology into these models is a crucial next step in better understanding the nature of cooperation.

Acknowledgments

We thank Gregor Caregnato for testing the participants, Natalie Ebner for allowing us to use her face photos, Sebastian Scholz for translating our NetLogo code into Pascal, and Mario Fific, Henrik Olsson, and Max Wolf for comments on an early version of the

manuscript. Funding for the project was provided by the Max Planck Society. This project was approved by the Max Planck Institute for Human Development Ethics Commission.

References

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, *111*(4), 1036–1060. doi:10.1037/0033-295X.111.4.1036

Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science*, *211*(4489), 1390–1396. doi:10.1126/science.7466396

Boyd, R. (1989). Mistakes allow evolutionary stability in the repeated prisoner's dilemma game. *Journal of Theoretical Biology*, *136*(1), 47–56. doi:10.1016/S0022-5193(89)80188-2

Chongsuvivatwong, V. (2010). *epicalc: Epidemiological calculator.* (R package version 2.11.1.0)

Cosmides, L., & Tooby, J. (1989). Evolutionary psychology and the generation of culture: II. Case study: a computational theory of social exchange. *Ethology and Sociobiology*, *10*(1-3), 51-97. doi:10.1016/0162-3095(89)90013-7

Ebner, N. C. (2008). Age of face matters: age-group differences in ratings of young and old faces. *Behavior Research Methods*, *40*(1), 130–136. doi:10.3758/BRM.40.1.130

Estes, W. K. (1955). Statistical theory of spontaneous recovery and regression. *Psychological Review*, *62*(3), 145–154. doi:10.1037/h0048509

Furlong, E. E., & Opfer, J. E. (2009). Cognitive constraints on how economic rewards affect cooperation. *Psychological Science*, *20*(1), 11-16. doi:10.1111/j.1467-9280.2008.02244.x

Hammerstein, P. (2003). Why is reciprocity so rare in social animals? A protestant appeal. In P. Hammerstein (Ed.), *Genetic and cultural evolution of cooperation* (pp. 83–94). Cambridge, MA: MIT Press.

Harrell, F. E. (2010). *Hmisc: Harrell miscellaneous.* (R package version 3.8-3)

Hill, R. A., & Dunbar, R. I. M. (2003). Social network size in humans. *Human Nature*, *14*(1), 53–72. doi:10.1007/s12110-003-1016-y

Kraines, D., & Kraines, V. (1989). Pavlov and the prisoner's dilemma. *Theory and Decision*, *26*(1), 47–79. doi:10.1007/BF00134056

Leeuw, J. de. (2001). *Reproducible research: the bottom line* (Tech. Rep.). Los Angeles: UCLA.

Leisch, F. (2002). Sweave: dynamic generation of statistical reports using literate data analysis. In W. Härdle & B. Rönz (Eds.), *Compstat 2002—Proceedings in Computational Statistics* (pp. 575–580). Physica Verlag, Heidelberg.

Linster, B. G. (1992). Evolutionary stability in the infinitely repeated Prisoners' Dilemma played by two-state Moore machines. *Southern Economic Journal*, *58*(4), 880–903.

Maynard Smith, J. (1982). *Evolution and the theory of games.* Cambridge, UK: Cambridge University Press.

Mealey, L., Daood, C., & Krage, M. (1996). Enhanced memory for faces of cheaters. *Ethology and Sociobiology*, *17*(2), 119-128. doi:10.1016/0162-3095(95)00131-X

Milinski, M., & Wedekind, C. (1998). Working memory constrains human cooperation in the Prisoner's Dilemma. *Proceedings of the National Academy of Sciences (USA)*, *95*(23), 13755–13758.

Molander, P. (1985). The optimal level of generosity in a selfish, uncertain environment. *Journal of Conflict Resolution*, *29*(4), 611-618. doi:10.1177/0022002785029004004

Nowak, M. A. (2006). Five rules for the evolution of cooperation. *Science*, *314*(5805), 1560–1563. doi:10.1126/science.1133755

Nowak, M. A., & Sigmund, K. (1992). Tit-for-tat in heterogeneous populations. *Nature*, *355*(6357), 250–253. doi:10.1038/355250a0

Pachur, T., Schooler, L. J., & Stevens, J. R. (in press). When will we meet again? Regularities in social contact dynamics reflected in memory and decision making. In R. Hertwig, U. Hoffrage,

& ABC Research Group (Eds.), *Simple heuristics in a social world.* Oxford: Oxford University Press.

R Development Core Team. (2010). *R: A language and environment for statistical computing.* Vienna, Austria. (ISBN 3-900051-07-0)

Rankin, D. J., & Eggimann, F. (2009). The evolution of judgement bias in indirect reciprocity. *Proceedings of the Royal Society of London, Series B*, *276*(1660), 1339-1345. doi:10.1098/rspb.2008.1715

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

Rieskamp, J., & Todd, P. (2006). The evolution of cooperative strategies for asymmetric social interactions. *Theory and Decision*, *60*(1), 69–111. doi:10.1007/s11238-005-6014-6

Roberts, G. (2008). Evolution of direct and indirect reciprocity. *Proceedings of the Royal Society of London, Series B*, *275*(1631), 173–179. doi:10.1098/rspb.2007.1134

Rudolph, U., Bohm, R., & Lummer, M. (2007). Ein Vorname sagt mehr als 1000 Worte: zur sozialen Wahrnehmung von Vornamen. *Zeitschrift für Sozialpsychologie*, *38*(1), 17-31. doi:10.1024/0044-3514.38.1.17

Sarkar, D. (2008). *Lattice: Multivariate data visualization with R.* New York: Springer. (ISBN 978-0-387-75968-5)

Schneider, W., Eschmann, A., & Zuccolotto, A. (2002). *E-prime reference guide.* Pittsburgh: Psychology Software Tools, Inc.

Stephens, D. W., McLinn, C. M., & Stevens, J. R. (2002). Discounting and reciprocity in an Iterated Prisoner's Dilemma. *Science*, *298*(5601), 2216–2218. doi:10.1126/science.1078498

Stephens, D. W., Nishimura, K., & Toyer, K. B. (1995). Error and discounting in the iterated prisoner's dilemma. *Journal of Theoretical Biology*, *176*(4), 457–469. doi:10.1006/jtbi.1995.0213

Stevens, J. R., Cushman, F. A., & Hauser, M. D. (2005). Evolving the psychological mechanisms for cooperation. *Annual Review of Ecology, Evolution, and Systematics*, *36*, 499–518. doi:10.1146/annurev.ecolsys.36.113004.083814

Stevens, J. R., & Hauser, M. D. (2004). Why be nice? Psychological constraints on the evolution of cooperation. *Trends in Cognitive Sciences*, *8*(2), 60–65. doi:10.1016/j.tics.2003.12.003

Todd, P. M., & Gigerenzer, G. (2007). Mechanisms of ecological rationality: heuristics and environments that make us smart. In R. I. M. Dunbar & L. Barrett (Eds.), *The Oxford handbook of evolutionary psychology* (p. 197-210). Oxford: Oxford University Press.

Wickelgren, W. A. (1974). Single-trace fragility theory of memory dynamics. *Memory and Cognition*, *2*(4), 775-780.

Winkler, I., Jonas, K., & Rudolph, U. (2008). On the usefulness of memory skills in social interactions: modifying the iterated prisoner's dilemma. *Journal of Conflict Resolution*, *52*(3), 375–384. doi:10.1177/0022002707312606

Wixted, J. T., & Carpenter, S. K. (2007). The Wickelgren power law and the Ebbinghaus savings function. *Psychological Science*, *18*(2), 133–134. doi:10.1111/j.1467-9280.2007.01862.x

Wu, J., & Axelrod, R. (1995). How to cope with noise in the Iterated Prisoner's Dilemma. *Journal of Conflict Resolution*, *39*(1), 183–183. doi:10.1177/0022002795039001008

Table 1: Experimental conditions

Condition #	Partners	Interactions	Replicates	Total trials
1	5	5	6	150
2	5	10	3	150
3	5	15	2	150
4	10	5	3	150
5	10	10	2	200
6	10	15	1	150
7	15	5	2	150
8	15	10	1	150
9	15	15	1	225

Table 2: Prisoner's Dilemma matrix

		Against:	
		Cooperate	Defect
Payoff to:	Cooperate	$R = 3$	$S = 0$
	Defect	$T = 5$	$P = 1$

Table 3: Strategy descriptions

Strategy	**Description (with `computer implementation` and *game theoretical definition*)**
ALLC (All Cooperate)	Always cooperate. *Probability of cooperating following T, R, P, S = (1, 1, 1, 1)*
ALLD (All Defect)	Always defect. *Probability of cooperating following T, R, P, S = (0, 0, 0, 0)*
CTFT (Contrite TFT)	Cooperate in the first round, then copy partner's choice in previous round. If agent mistakenly defects, switch to cooperating. `if this is first interaction with partner, cooperate` `if partner cooperated in previous move, cooperate` `if partner defected in previous move & this is your second or third interaction with partner, defect` `if partner defected in previous move & this is your fourth or more interaction with partner, look at own move before previous move:` `if you cooperated, defect` `if you defected, look at partner's second previous move:` `if partner cooperated, cooperate` `if partner defected, defect` *No game theoretical definition—memory-2 strategy*
GRIM (Grim Trigger or Friedman)	Cooperate until partner defects, then always defect. `if this is first interaction with partner, cooperate` `if partner defected in previous round, defect` `if partner cooperated in previous round, look at own previous move:` `if you cooperated, cooperate` `if you defected, defect` *Probability of cooperating following T, R, P, S = (0, 1, 0, 0)*
GTFT (Generous TFT)	Cooperate in the first round, then copy partner's choice in previous round. If partner defected, cooperate with probability 0.33. `if this is first interaction with partner, cooperate` `if partner cooperated in previous round, cooperate` `if partner defected in previous round, defect with probability 0.66` *Probability of cooperating following T, R, P, S = (1, 1, 0.33, 0.33)*
RAND (Random)	Randomly choose to cooperate or defect for each round. *Probability of cooperating following T, R, P, S = (0.5, 0.5, 0.5, 0.5)*
TFT (Tit-for-tat)	Cooperate in the first round, then copy partner's choice in previous round. `if this is first interaction with partner, cooperate` `if partner defected in previous round, defect` `if partner cooperated in previous round, cooperate` *Probability of cooperating following T, R, P, S = (1, 1, 0, 0)*
TF2T (Tit-for-two-tats)	Cooperate in the first two rounds, then copy partner's choice in previous round. If partner defected, look back another round, and if partner defected then, defect, otherwise cooperate. `if this is first interaction with partner, cooperate` `if partner cooperated in previous round, cooperate` `if partner defected in previous round & this is your second interaction with partner, cooperate` `if partner defected in previous round & this is your third or more interaction with partner, look at round before:` `if partner cooperated, cooperate` `if partner defected, defect` *No game theoretical definition—memory-2 strategy*
WSLS (Win-stay, Lose-shift or Pavlov)	Cooperate following mutual cooperation or mutual defection, otherwise defect. `if this is first interaction with partner, cooperate` `if you cooperated and partner cooperated, cooperate` `if you defected and partner defected, cooperate` `if you cooperated and partner defected, defect` `if you defected and partner cooperated, defect` *Probability of cooperating following T, R, P, S = (0, 1, 1, 0)*

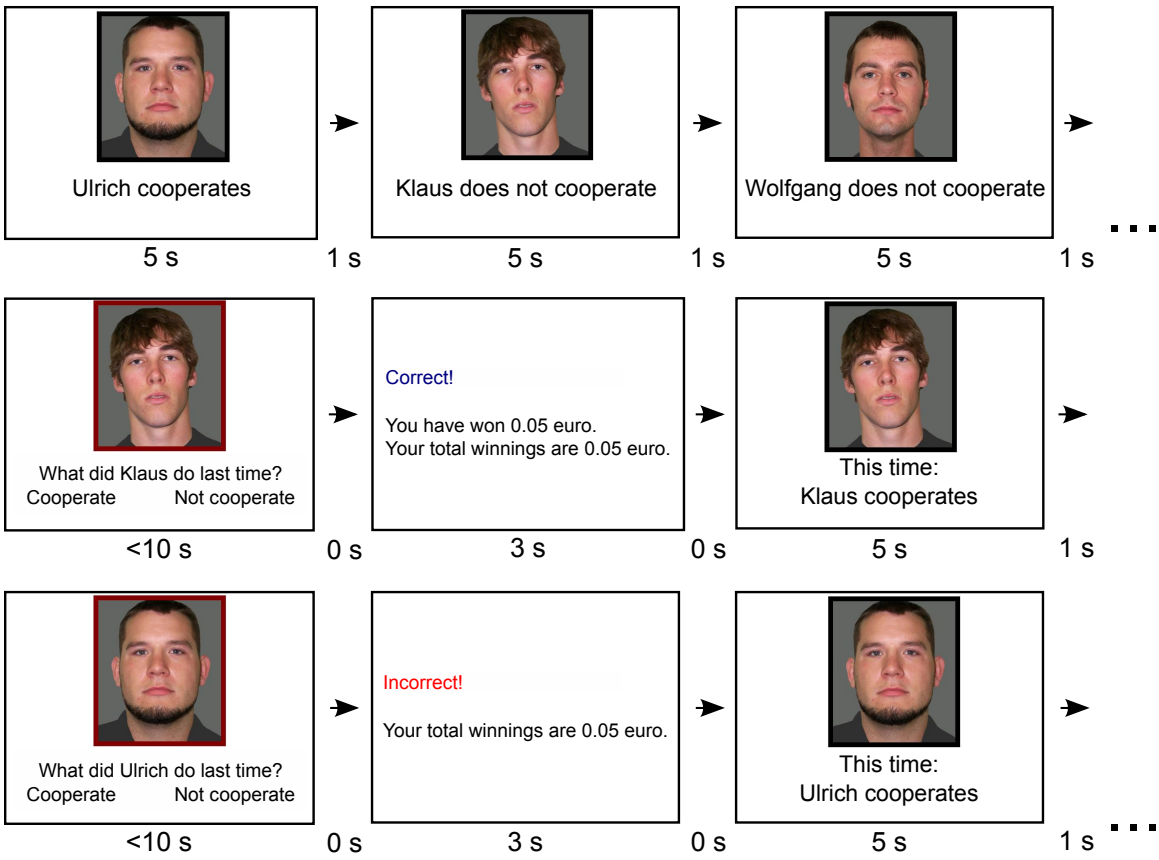

Figure 1. Screen shots of the cooperative memory task. In the first round of the task (top row), participants observed an image and the name of each partner, along with the current action. After viewing this for each partner, participants were asked for a partner's previous choice, given feedback on his or her response, and updated on the partner's new choice before moving on to the next partner (middle and bottom rows). Numbers below screens give presentation times for screens and between screens.

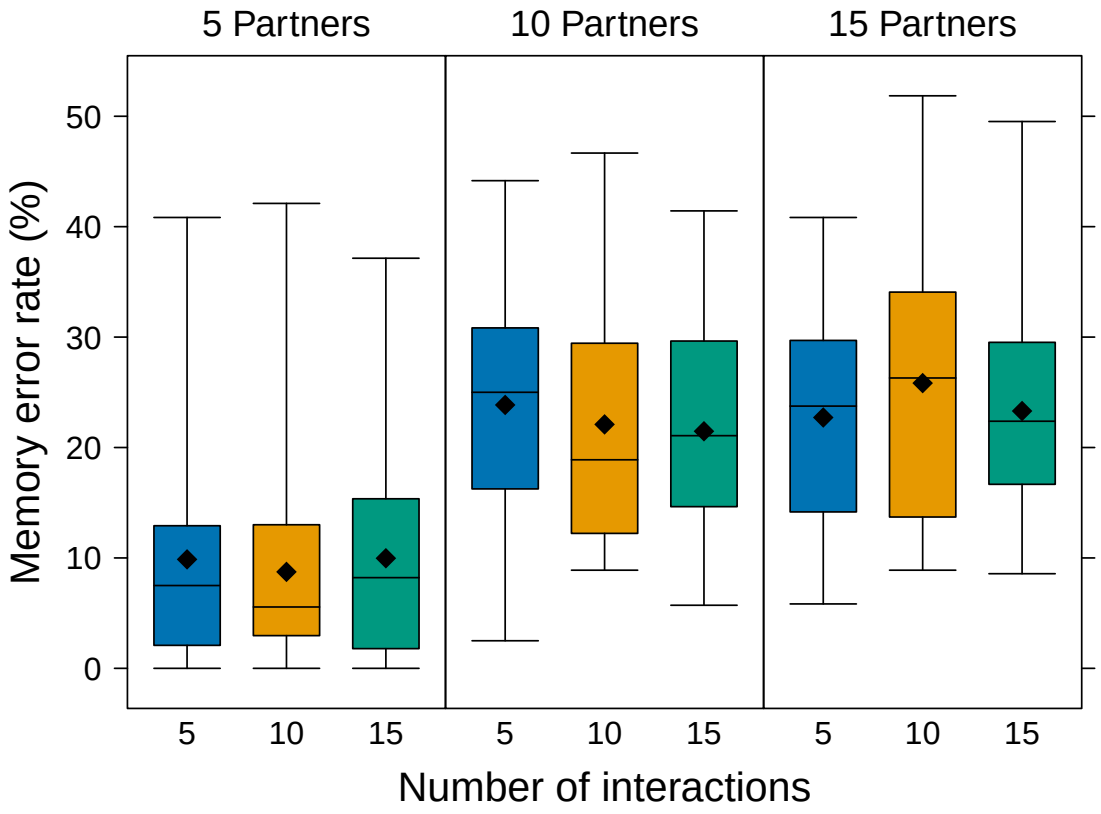

Figure 2. Memory error rate as a function of partner number and number of interactions. Boxplots show that the error rate increased with group size ($N = 24$ participants in each of nine conditions). The number of interactions per partner, however, did not influence error rate. Diamonds represent the mean, lines represent the median, boxes represent the interquartile range, and whiskers represent 1.5 times the interquartile range.

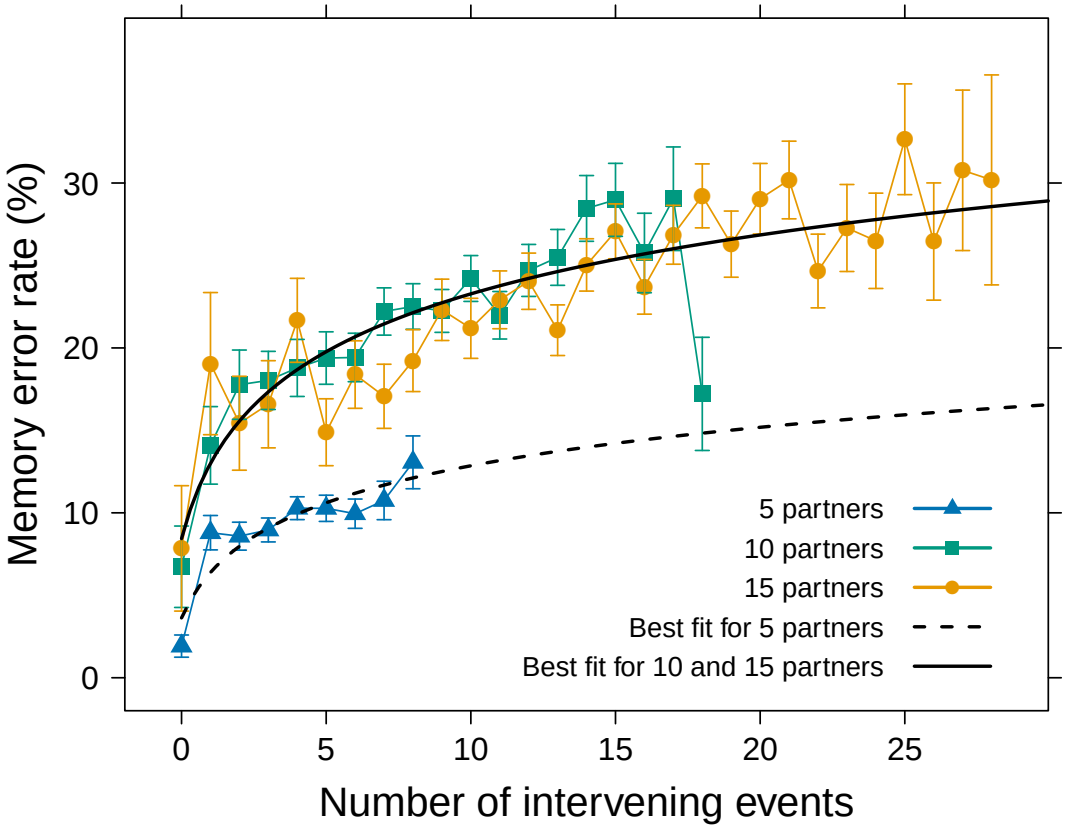

Figure 3. Interference effects on memory accuracy. The mean (±SEM) error rate increased with more intervening interactions across all three group sizes (collapsing across the number of interactions per partner), with the effect more pronounced in group sizes of 10 or 15. The smooth lines represent the least-squares best-fit Wickelgren's (1974) power function of memory to either the 5-partner data or combined 10- and 15-partner data (for both lines, $R^2 = 0.90$).

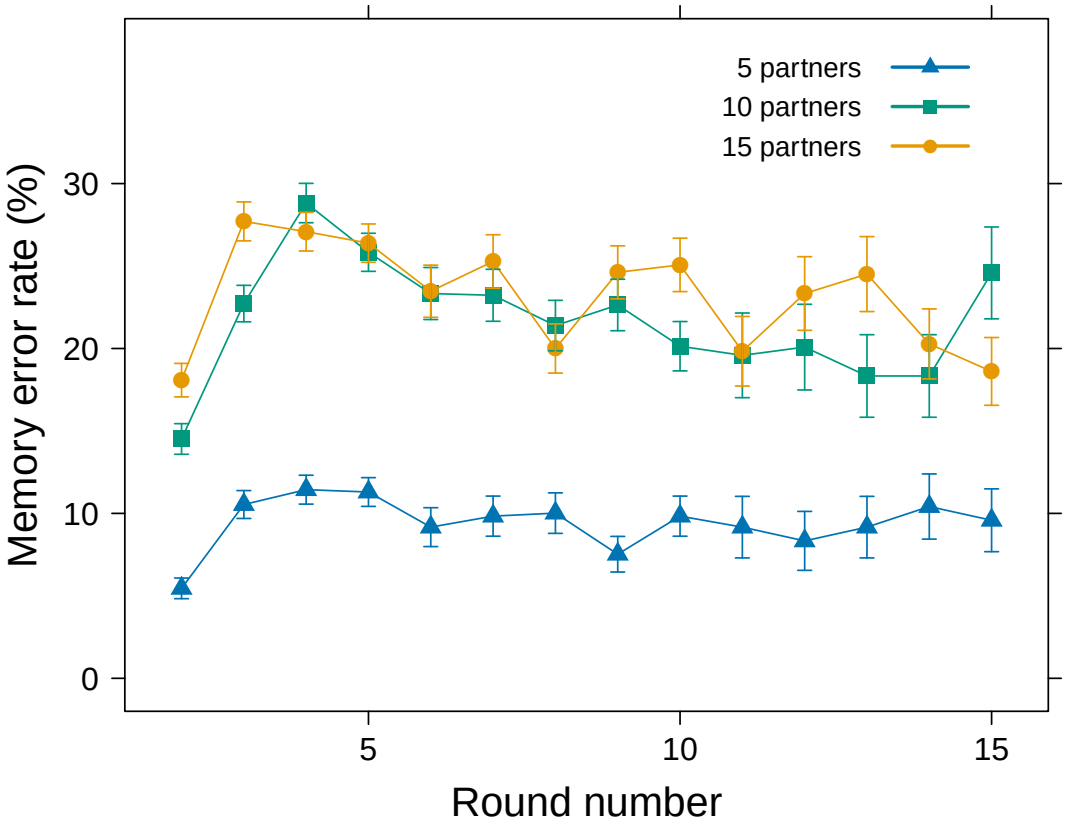

Figure 4. Error rate as a function of round number. The mean (±SEM) error rate increased in the first three or four rounds before decreasing.

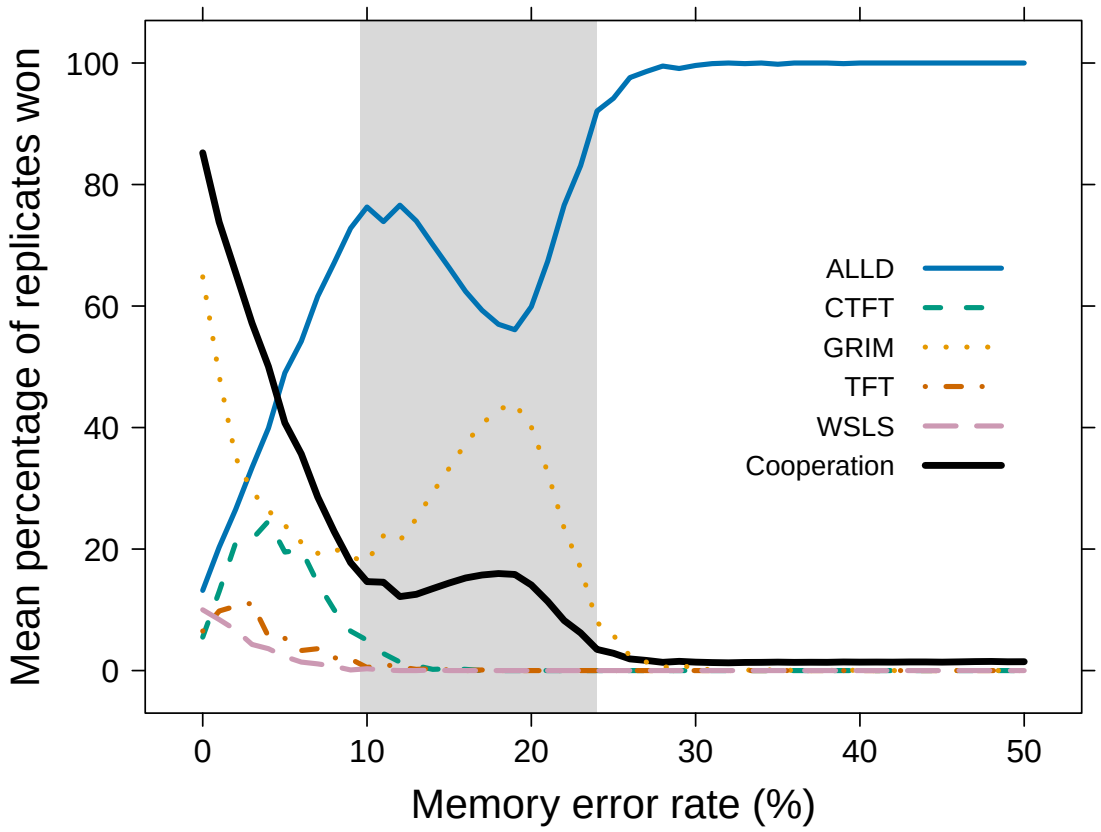

Figure 5. Agent-based simulations of error rate effects. When varying error rates across a range of values, GRIM, CTFT, TFT, WSLS, and ALLD survived with few errors (we do not show strategies with success rates lower than 0.05%). At higher rates (e.g., error rates observed in the experiment are shaded), however, ALLD and GRIM outperformed the other strategies. The proportion of cooperative choices made by all agents in the last generation decreased rapidly with increasing error rate.

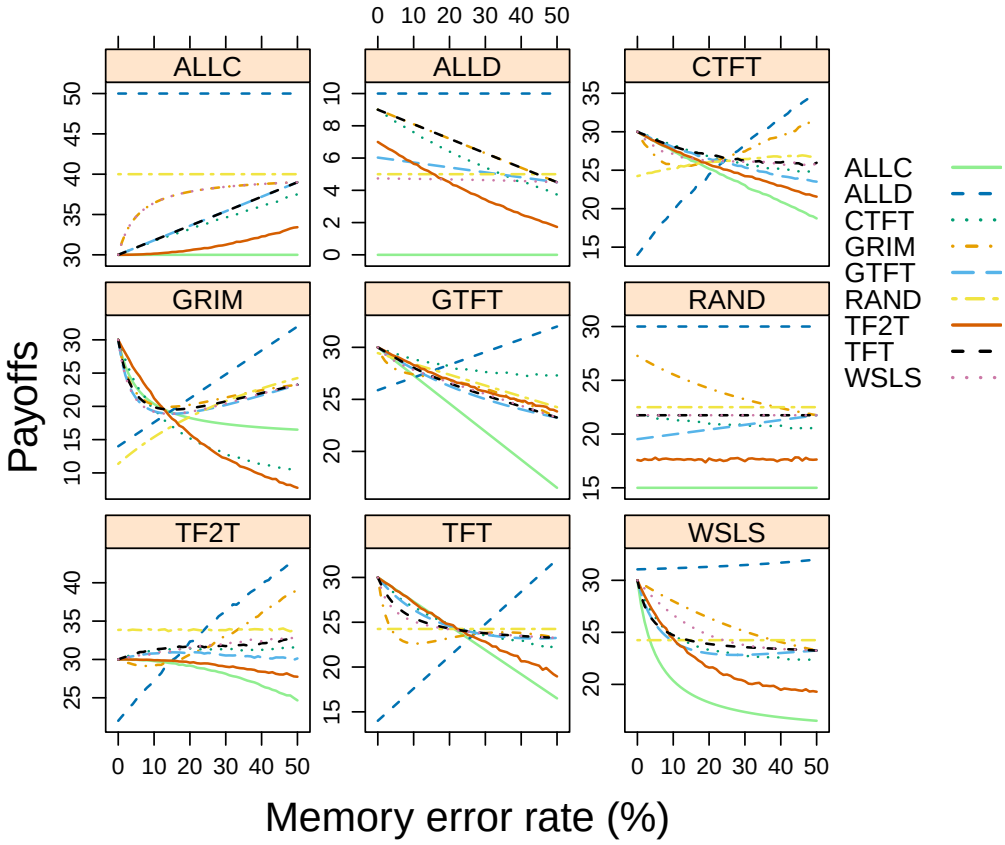

Figure 6. Game-theoretical payoffs of strategies as a function of error rate. For each strategy, we calculated how all strategies perform against that strategy over a range of error rates. When the strategy playing against itself has a higher payoff than any other strategy playing against it, this is an evolutionarily stable strategy (ESS). Strategies CTFT and TF2T were simulated rather than analytically calculated.

Appendix A
Participant Instructions

Below is a translation from German of the participant instructions.
Instructions

In this experiment, you will repeatedly interact with a number of hypothetical partners. For each interaction, your partner will choose either to *cooperate* or *not cooperate*. **Your task is to recall the last action for each partner**.

To give you a concrete example of what this might mean, imagine that you repeatedly go out to dinner with each partner. At the end of the meal, you each must decide *individually* whether to contribute to a tip for the waiter. If your partner tips, this would be an instance of cooperating, but if your partner does not contribute to the tip, then this is not cooperating.

In this task, we will assess how well you remember whether each partner cooperated or not *the last time you interacted*.

Procedure:
First you will be shown for each partner whether he/she cooperates or not. You should try to remember each partner's action. In the example below, Natalie cooperates.

After observing all of the partners' actions one after the other, it follows the retrieval of the actions of the individual partners. For this purpose you will meet each partner again but not necessarily in the same order as in the beginning. Each time you will be asked whether the displayed partner cooperated or not the last time that you interacted with him/her.

Press 'k' for 'cooperate' or 'n' for 'not cooperate'. You will have ten seconds to respond. If you wait longer than ten seconds, the question will be skipped.

After each response, you will learn whether you were correct. Thereafter you will see what the partner decides to do this time. In the example below, Natalie doesn't cooperate this time.

This is now the action that you should try to keep in mind. **The task always is to recall the last action for the partner**. Then there will follow the retrieval, feedback and new action for the next partner and so on.

Please respond as accurately as possible. You will receive 5 cents for every correct response (in addition to your show-up fee of 5 euros).

Altogether you can receive an additional payment of 8 euros on average. For incorrect responses or skipped questions, you will receive no payment.

Generally:
For this experiment your partners will be grouped, such that you will repeatedly interact with the same partners before moving on to a new group of partners. Each group will have a different number of partners which you will interact with a different number of times. After you complete a group, you can have a short break before beginning the next group. The whole task should last about 1,5 hours.

You will begin with a practice phase in which you can see how the task works without earning money. If you have any questions, please ask the experimenter. If you are ready to begin the practice phase, please press <space bar> on the computer keyboard.

Appendix B
Participant Questionnaire

Below is a translation from German of the participant questionnaire.

1. Do you know one/some of the depicted persons?

2. Did you associate memories of a/some certain person/s with one/some of the used names?

3. Of 10 decisions that you made how often did you guess on average?

4. Of 10 of your partner's actions how often, you think, did the interaction partners cooperate on average?

5. Did you pursue a certain strategy for memorizing the partner's actions? If you did, please describe the strategy you used.

6. What did you do when you could not remember the action from the previous round?

7. Do you have comments or suggestions?