A new consequence of Simpson's paradox: Stable co-operation in one-shot Prisoner's Dilemma from populations of individualistic learning agents

Nick Chater[1*], Ivo Vlaev[1], Maurice Grinberg[2]

[1] Department of Psychology

University College London

London, UK

[2] Central and East European Center for Cognitive Science

New Bulgarian University

Sofia, Bulgaria

n.chater@ucl.ac.uk

i.vlaev@ucl.ac.uk

mgrinberg@nbu.bg

* corresponding author

Abstract

Normative theories of individual choice in economics typically assume that interacting agents should each act individualistically: i.e., they should maximize their own utility function. Specifically, game theory proposes that interaction should be governed by Nash equilibria. Computationally limited agents (whether artificial, animal or human) may not, however, have the capacity to carry out the sophisticated reasoning to converge directly on Nash equilibria. Nonetheless it is often assumed that Nash equilibria will be obtained, in any case, if agents embody simple learning algorithms like reinforcement learning. If so, then learners should converge on Nash equilibria, after sufficient practice in playing a game---and hence, for example, individualistic agents should end up playing *D* (defect) in one-shot Prisoners' Dilemmas (PD). In an experiment and in a multi-agent simulation, we show, however, that this is not always the case---under certain circumstances, reinforcement learners can converge on co-operative behaviour in PD. That is, even though each agent would receive higher pay-off from switching to *D*, agents obtain more reinforcement, on average, from playing *C*, and hence *C* is more strongly reinforced. This effect arises from a well-known statistical paradox, Simpson's paradox. We speculate that this effect may be relevant to some aspects of real-world human co-operative behaviour.

Introduction

Modern society rests on the fact that people are routinely kind to strangers, whom they will not meet again. People take turns in shops; give way to other drivers; risk their lives attempting to save drowning swimmers.

Much apparent altruism can be "explained away" in many situations, by observing the existence of hidden benefits: by reproductive advantages for "selfish genes" (Dawkins, 1976) when interacting with kin (Hamilton, 1964); by the long view, if our kindness may be reciprocated (Axelrod & Hamilton, 1981; Axelrod, 1984; Nowak & Sigmund, 1989); or by social approval, if we are being observed by peers (Wedekind & Braithwaite, 2002; Satow, 1975). Pure kindness may, therefore, arise by overextension of "impure" kindness to cases where hidden benefits are not available.

Pure kindness remains puzzling, however, in two respects. The first puzzle is stability: why is this kind of kindness not unlearned through years of social behaviour. Why do people not learn that, when interacting with people we will never see again, selfishness pays (Schneider & Mockus; 1974; Cooper, DeJong, Forsythe, & Ross, 1996). The second puzzle is the folk belief in "good Karma:" the view that, independent of direct causal links (e.g., reciprocation) that "what goes around comes around," that "kindness begets kindness." That is, not only do we persist in being kind to strangers for no pay-back; but we form an illusory belief that it is, nonetheless, good for us in terms of long-run self-interest (Baron, 1997).

Ethical principles surely explain some aspects of pure kindness. Yet, we suggest, another force may be at work---that even when kindness does not *cause* good outcomes, it may be *correlated* with good pay-offs. Correlation with good outcomes, whether underpinned by causation or not, determines whether behaviours is repeated, according to widely used theories of associative learning, derived from Thorndike and Skinner (see Sutton & Barto, 1998, for a review).

Why might such non-causal correlations between pure kindness and good outcomes occur? If people engaged in pure kindness randomly, they would not. But we know that this kindness is not random, but is influenced by a range of "situational" factors (e.g., Colman, 1995). For example, kindness, like other social behaviors, is substantially influenced by whether agent and recipient identify with each other (e.g., are in the same national, professional, religious, or other, group); whether they make eye-contact; whether the environment primes helping behavior (e.g., church vs. casino); tendency to imitate recently observed behavior of others (Hurley & Chater, 2005). Factors such as these lead, we suggest, to pure kindness being highly correlated across individuals: in some situations, people will tend to show high levels of pure kindness, in others they will not.

This correlation means that people tend to be kind to others in those situations in which people tend to be kind to them, even where the two are not causally linked. Optimal pursuit of self-interest, as predicted by evolutionary and economic models, would predict that people should take the benefits of kindness, but give out none themselves, rapidly eliminating pure kindness. Yet, if behaviour is shaped by simple principles of reinforcement learning, the correlation between own kindness and good outcomes from others' kindness could lead to kindness being reinforced, entrenching pure kindness.

We propose that one factor contributing to these effects arise from a well-attested cognitive blindspot—that people, and their associative learning mechanisms, cannot correct for biased samples (see Fiedler, 2000; Fiedler & Juslin, 2006). Specifically, people's experience of cooperation may constitute a biased sample, for which they are unable to correct. The core idea is that, however co-operative people may be, there will some situations which especially encourage co-operation (e.g., ordering queuing in shopping); we call these, *nice* situations; and there are some situations where people do not tend to co-operate (e.g., barging ahead at the opening of the January sales); call these *nasty* situations. Because all players are influenced by whether they are in nice or nasty situation, their behaviour will be correlated. Overall, people will tend both to cooperate, and be cooperated with, in nice situations; but the reverse in nasty situation. Note that this looks superficially like reciprocation---when I am (un)cooperative, people are (un)cooperative with me. But there is no causal connection between each person's action—they are merely correlated because they are both influenced by the situation. We shall see that, under certain circumstances of this kind, it is possible for this correlation to have the result that the average reinforcement obtained for playing *C* is greater than the average reinforcement for playing *D*. Hence, although naked self-interest is actually best served by continual defection against all-comers, a population of self-interested learners unable to correct for sampling bias may nonetheless end up cooperating. Indeed, reflecting on the pay-offs of both outcomes, such learners might be expected to erroneously conclude that, because own cooperation is correlated with the cooperation of others that 'what goes around comes around.' Here, we shall focus on providing an "existence proof" of this phenomenon---first by simulations of interacting reinforcement learning agents, and second by an experimental demonstration with human participants.

Normative theories of interactive decision making

In economics, game theory has typically been viewed as providing the normative theory of interacting agents (Fudenberg & Tirole, 1996). In particular, it is typically assumed that interacting agents should follow Nash equilibria---i.e., patterns of behaviour in which each agent cannot improve its outcome, given the other agent's behaviour. This type of normative explanation is fundamentally individualistic---that is, each agent chooses its behaviour to maximize the outcome according to its own preferences (although this does not preclude, of course, that an agent's preferences may be altruistic, rather than selfish).

Three issues concerning the status of this style of normative explanation may be raised, which have relevance to understanding interacting systems of agents, whether biological or artificial.

The first issue is that many scenarios involving interaction define games that have many Nash equilibria, only some of which are observed. A substantial body of game-theoretic research attempts to refine the notion of Nash equilibrium (e.g., by introducing notions such as the "trembling hand" equilibrium), to attempt to pick out those equilibria that are experimentally observed.

The second issue concerns the opposite problem---not that there are too many Nash equilibria; but that observed behaviour (particularly in real-world animal or human behaviour; or in the laboratory) does not appear to correspond to a Nash equilibrium at all. To focus on the most celebrated examples, which we will discuss extensively below, people frequently co-operate, i.e., play *C*, in scenarios which appear to have a one-shot Prisoners' Dilemma structure (see e.g. Dawes & Thaler, 1988). But the only Nash equilibrium in the Prisoners' Dilemma is that people play *D* (see Figure 1). Whatever agent 2 does, agent 1's pay-off will be higher if it plays *D*; and vice versa. The mismatch between normative theory and empirical results concerning human interactive behaviour can be addressed in a number of ways: e.g., by assuming that if people play the same game many times, against different agents, they will eventually converge on the Nash equilibrium; that there are additional pay-offs that affect the agents' utilities, in addition to those specified in the game; or that a revised normative theory that switches from an individualist, to a 'team' perspective, may be appropriate (Bacharach, 1999; Gold, in press; Hurley, 1989). Yet a further line of explanation, concerning reciprocation, arises for a case that we do not consider: that of repeated PD against the same opponent (and where the identity of agents is known). Repeated PD is, from a normative game-theoretic standpoint, a completely different game, because each trial of the game can potentially influence the other agent's play in the next trial. Repeated PD is best modelled in terms of agents choosing strategies which define their sequence of responses to other agents' strategies over the course of repeated interaction. Conventional game theory has little purchase on this problem---in "strategy space" the repeated PD game has many possible Nash equilibria, one of which is*C* (by the "folk-theorem," e.g., Kreps, 1990). Repeated PD has been extensively studied using agent-based simulations (e.g., Axelrod, 1984). We mention repeated PD here principally to distinguish it from the one-shot PDs that we consider here. In one-shot PD, players are anonymous. One-shot PDs can, however, be played many times. Crucially, players are paired afresh with new partners for each game (and even if, in a simulation or experimental context, a partner in a later game is the same as the partner in an earlier game, the player will not be aware of this). We shall consider one-shot PDs only henceforth.

INSERT FIGURE 1 ABOUT HERE

Alongside the issues of too many Nash equilibria, and patterns of behaviour that do not seem to fit any Nash equilibrium, stands a third issue. This concerns the underlying computational machinery that agents require for the Nash equilibrium to be attained---machinery that may appear to substantially exceed the capacities of human reasoning. From the normative perspective of game theory, the stability of the Nash equilibrium rests on the assumption that each agent adopts 'best-reply' reasoning. That is, the agent attempts to make the 'best reply' to whatever the other agent will do; and, moreover, assumes that the other agent(s) will also follow best-reply. Best-reply reasoning rules out patterns which are not Nash equilibria, because, by definition, outside a Nash equilibrium, one of other agent can make a better reply, in terms of its own utility function---and this is precisely what the conditions of a Nash equilibrium rules out. Yet best-reply reasoning is typically extremely subtle---involving the ability to reason recursively about the behaviour of the other. This may be difficult for cognitively bounded agents (indeed, human game playing seems to be subject to extremely severe bounds, in terms of the depth of such recursive inference, Colman, 2003). This might appear to suggest that the normative predictions of game theory may not be helpful in understanding the behaviour of artificial and biological agents.

One natural way to address this last issue, of computational limitations, is to assume that, although agents may be computationally very simple, elementary processes of learning may lead them to fall into Nash equilibria. There has, in particular, been considerable interest in research on behavioural models game-playing on reinforcement learning (e.g., Erev & Roth, 1998): i.e., that agents tend to repeat behaviours according to the average degree of "reinforcement" with which they are associated---i.e., the average utility for the agent of the outcome of the game associated with each response. From a psychological point of view, this corresponds to following Thorndike's (1911) classic "Law of Effect"---repeating behaviours to degree that they are followed by positive outcomes; and stamping out behaviours to the degree that they are followed by negative outcomes. There is hope that this tack may also help deal with the first two issues, concerning equilibria. Perhaps agents will only stabilize of empirically observed Nash equilibria and not those that are not observed (e.g., Binmore, 1994). Moreover, perhaps (as we shall see below) reinforcement learners may under certain circumstances fail to converge on Nash equilibria---and this might explain empirical observed violations of the Nash equilibrium.

Agents that learn by reinforcement are attractive in a number of ways. They learn according to very simple principles; they involve no "reasoning" (best-reply or otherwise) about the other agents behaviour; and they appear likely to fall naturally into Nash equilibria. Simplifying somewhat, the idea is that if a group of agents falls into a stable pattern of behaviour at all (of course, it is possible that behaviours might cycle or move ergodically, without settling), then the agents must be in a Nash equilibrium. Concretely, both agents should end up playing *D* in Prisoners' Dilemma. The intuition is that, whatever the other agent does, each agent will obtain more reward from playing *D* rather than *C*; and hence a reinforcement learning algorithm will gradually eliminate *C* in favour of *D*.

As we show in this article, however, reinforcement learners need not, despite this apparently ineluctable line of argument, converge on Nash equilibria. This is because reinforcement learning methods are based on the *average* amount of reinforcement that each behaviour actually receives; and it is possible that these averages can be systematically misleading, if the two responses, *C* and *D*, are not generated independently, but are correlated by other aspects of the game that can affect the outcome. The formal basis of this effect is a statistical phenomenon known as Simpson's paradox (Simpson, 1951).

In this article, we show that, against intuition, a simple reinforcement learning algorithm can lead to stable non-Nash-equilibria *CC* responses in multiple one-shot PD games, against different agents. We also show that this result can occur human participants, where the typical drift toward playing *D* with practice is replaced by stable *CC* responding. In the *Discussion*, we ask whether, in practice, the Simpson's paradox effect that we describe may be one force that helps ensure the stability of cooperative behaviour in real world human interactive behaviour. We also suggest that our results throw a new light of the role of game theory as a normative framework for understanding the patterns of behaviour of interacting agents.

Simpson's paradox: How reinforcement learners find stable non-Nash equilibria

Must populations of reinforcement learners converge on DD?

As we have seen, from the definition of PD, it follows that an agent should, from a purely individualistic perspective, play *D*, whether the other agent plays *C* or *D*. To spell this out, from the point of view of, say, Agent 1, there are two ways the world can be. The world may be such that Agent 2 has played *C*; or the world may be such that Agent 2 has played *D*. The pay-off for *D* is greater than the pay-off for *C*, either both of these situations. It *seems* immediately to follow that the pay-off for *D*, averaged across these two states of the world, must be greater than the average for *C*, averaged across those two states of the world. Therefore, it *seems* to follow that the average reinforcement to *D* must necessarily be greater than the average reinforcement to *C*, and hence that a reinforcement learner will reinforce *D*, and eliminate *C*; and hence that a set of interacting reinforcement learners, playing one-shot prisoner's dilemmas against different players, must attain the Nash equilibrium of *D*, *D*.

But this chain of argument, although appealing, is flawed, because, despite intuition, it turns out that the *average* outcome for *C* can be greater than *D*, even though the outcome for *D* is higher than *C* in both of the two possible states of the world---when the other agent plays *C* or the other agent plays *D*. Hence, it is possible that a reinforcement learner may not necessarily drift inexorably to *D*, but may continue to cooperate by playing *C*. This strange phenomenon is an example of Simpson's paradox, which we briefly describe.

Simpson's paradox and problem of biased sampling

INSERT FIGURE 2 ABOUT HERE

Simpson's paradox demonstrates that two variables can be negatively correlated, even though, given any fixed value of a third variable, they are positively related (or vice versa). To illustrate consider the set-up depicted in Figure 2. Suppose that we consider the introduction of a new inhaler to help people with breathing difficulties. The inhaler is typically prescribed for people with asthma---but it turns out also to be helpful in improving the breathing of people without asthma. Thus, whatever the value of the third variable, whether the person has asthma or not, the proportion of people with clear breathing is higher if the inhaler is used. Specifically, in the fictitious data in Figure 2, the inhaler increases the proportion of asthma sufferers with clear airways from 10% to 44%; and increases the proportion of non-asthma sufferers with clear airways from 90% to 99%. But if we average our patient data together, we get the opposite result! People who use the inhaler typically have less clear breathing than people who do not: overall, 50% of people using an inhaler have clear airways, compared with 82% who do not use the inhaler. Thus we have an apparently paradoxical result: that the inhaler appears to harm breathing overall, even though it helps the breathing of both the asthmatic and non-asthmatic groups considered separately. The explanation of the paradox is, of course, that the inhaler really does help both groups; but that inhaler users typically have more breathing problems, because inhaler-users typically have asthma. That is, the people using inhalers are not an unbiased sample of people in our population of patients---but rather, the fact that someone uses an inhaler is a clue that they have breathing problems. Therefore, overall, people who use inhalers may have worse breathing than those who do not, even if inhalers are beneficial both to people who do, and who do not, have breathing problems.

The application of Simpson's paradox in explaining how CC may be maintained theogh learning mechanisms is, as far as we know, novel. However, the phenomenon has been used, in a very different way, in attempts to explain how genes for cooperation can evolve, including in interactions such as Prisoners' Dilemma (Nowak & May, 1992; Sober & Wilson, 1998), although we shall not consider these arguments further here.

Simpson's paradox and how reinforcement learners can keep playing C

Suppose that, due to the influence of some external factor (we consider some specific possibilities below), agents initially tend to behave in a positively correlated manner. If so, then learning of one agent's decision to choose *C* or *D* will be informative about the other agent's behaviour. This is the type of biased sample in which Simpson's paradox can arise.

Specifically, if agents' responses are positively correlated, this means that *CC* and *DD* outcomes are the most common; and this implies further that when *D* is played, it is typically associated with a low outcome (the *DD* pay-off, in Figure 1, where both prisoners defect is low in the PD game). Occasionally, of course, the correlation is broken; and then the agent who plays *D* gets a high pay-off (Tempt); and the agent who plays *C* gets a low pay-off (Sucker). But assuming the correlation between responses is high, then low *DD* pay-off dominates---hence the average reinforcement for playing *D* is low. Conversely, where either agent plays *C*, the correlation ensures that, mostly, the pay-off is the reasonably high *CC* pay-off (this pay-off is, by definition of the PD, higher than outcome when both players defect, because *CC* > *DD*). Only on the occasions where the other agent plays *D* is a low pay-off received. Because of this correlation, *C* is associated on average with somewhat less than the good *CC* pay-off; and *D* is associated on average with somewhat more than the poor *DD* pay-off. Thus, it is possible for *C* to receive a higher average reinforcement than *D*, even though the agent will always benefit by playing *D*.

Thus, given a biased sample of *C* and *D* responses, the average pay-offs obtained by the reinforcement learner may be greater for *C*, because *C* is mostly associated with fairly good outcomes, in which the other agent co-operates; whereas *D* is most associated with poor outcomes, in which the other agent does not cooperate. Of course, the agent would do *even* better by consistently playing *D*; but a reinforcement learner is not able to exploit this---it is not able to notice and correct for biased sampling. Instead, its behaviour is driven by average pay-off, even when the sampling that leads to that average is systematically biased.

There is a good deal of empirical evidence that people also appear to be unable to correct effectively for sampling bias: (a) theories of associative learning, based on principles similar to reinforcement learning, compile averages, without correcting for sampling bias (Pearce & Bouton, 2000; Waldmann, 1996); (b) in direct experimental tests, people fall into Simpson's paradox (Curley & Browne, 2001; Fiedler, Walther, Freytag, & Nickel, 2003; Spellman, Price, & Logan, 2001); (c) a wide range of cognitive and social judgment biases can be explained by an inability to correct for biased sampling (e.g., Fiedler, 2000). We report, below, a direct empirical demonstration that co-operation may be high, and stable, where Simpson's paradox applies.

Inducing a positive correlation between agents' behaviour from game pay-offs

Before exploring this phenomenon in more detail, in simulations and human experiments, we consider one simple, and experimentally useful, way in which a correlation between agents' behaviour may arise. The case that we explore henceforth is that the structure of the game itself may induce such correlations. It is well-known that different types of PD lead to different degrees of co-operative behaviour (different proportion of *C* choices) (Rapoport & Chammah, 1965). Below, we assume that there are two kinds of game, *nice* and *nasty* games, which participants play in an intermixed fashion. Here, 'type of game' is the variable that induces the crucial correlation between the behaviour of the two agents. As we note in the *Discussion*, many other sources of such correlations may be of interest. For example, considering human social behaviour, it is frequently observed that co-operative behaviour is much more likely to occur if both people are within the same group, rather than in different groups (these groups can be defined by nationality, profession, or even arbitrarily, as "red" and "green" groups). If people must play PDs with an anonymous other player, knowing only whether they are from the same, or different group, then it is likely that they will cooperate more strongly with group members. Thus, within-group encounters will be overrepresented by *CC*; and between-group encounters will be associated by *DD*. This is just the correlation that is relevant for giving rise to Simpson's paradox. More generally, to the extent that people's (or animal's) choice behaviour in real decision making situations is influenced by external variables, which are relevant to both players in the game, then such positive correlations may arise.

Preview

We have tested these theoretical ideas in two ways. We present the results of an experiment in which people play repeated PDs. The second test is that we modelled the results of this experiment, using a population of simple reinforcement learners. As we shall see, both human agents and reinforcement learners can stably cooperate in PD, when Simpson's paradox applies.

Simpson's Paradox Meets Prisoner's Dilemma: An Experimental Demonstration

We gave participants a sequence of one-shot Prisoner's Dilemmas (Figure 1) in which agents were anonymous and randomly paired afresh for each trial. Situational factors that influence behaviour correspond to two different types of game: *nice* and *nasty*. The *nice* situation (Figure 3a) was a Nice PD where pay-offs gave little incentive for noncooperation; and by choosing the “kind” action 1, a player loses, e.g., one extra unit of reward, and thereby gives the other player ten reward units. The *nasty* situation (Figure 3b) was a Nasty PD, where there is substantial incentive for noncooperation, because by choosing the “kind” action 1, a player loses, e.g., five extra units of reward (instead of ending up with zero loss if choosing action 2), and thereby gives six reward units to the other player (who loses five instead of eleven units). In all cases, of course, standard game theory would predict that the self-interested agent should not cooperate, but should play defect. But Rapoport and Chammah (1965) show that changing the pay-offs, as indicated here, substantially modifies people’s propensity to cooperate (validating the “Nice” and “Nasty” labels above).

INSERT FIGURE 3 ABOUT HERE

Thus, we already have the starting point for the operation of Simpson’s paradox. If the other player plays *C*, the self-interested agent will have a higher pay-off by playing *D*; and if the other player plays *D*, the self-interested agent will have a higher pay-off by playing *D*. But because the other players move is *correlated* with the first players choice, it does not necessarily follow that the average pay-off for *C* will be less than the average pay-off for *D*. To see this intuitively, let us take the extreme case, that all players begin by cooperating with 100% reliability in the nice game; and defecting with 100% reliability in the nasty game. In the first case, as both players cooperate, they receive the *CC* outcome (see Figure 1). On the other hand, in the second case, where both players defect, they receive the less good *DD* outcome. Hence, defection leads to a lower average pay-off than cooperation. Suppose, moreover, that the players are not quite perfectly correlated---they occasionally play *D* in a nice game; or *C* in a nasty game. *If* they are able to conditionalize on the particular type of game that they are playing (nice vs. nasty), then the players should realize that the occasional

D is advantageous in the nice game, and hence inceasingly repeat it; and that the occasional *C* in a nasty game is disadvantageous, and eliminate it. Such learners would end up in playing *D* for all games, and following the Nash equilibrium. Yet, if the learners, instead, average over all instances where they play *C* and *D*, independent of whether the game is nice or nasty, then they may find that the average pay-off for *C* (Co-op with an occasional Sucker, see Figure 1) is higher than for *D* (Defect with occasional Tempt); and hence their level of cooperation might be stable or even increase.

To get the cleanest possible instance of this phenomenon, we added a further factor: we make the overall levels of pay-off in "nice" games higher than those in "nasty" games. In game theory, where each individual game is considered independently, and strategy is invariant over additive changes to all pay-offs, this should make no difference. But for reinforcement learners who do not conditionalize over game types, this will strengthen Simpson's paradox still further---because the *CC* pay-off in the nice game will be especially high in relation to the *DD* pay-off in the nasty game. Specifically, we implemented this by making the outcomes of the 'nice' games positive; and the outcomes of the 'nasty' games negative.

The set-up used here is, of course, deliberately quite extreme: our goal was to present only an "existence proof" of a very unexpected phenomenon – that reinforcement learners can stably play *C*, in a sequence of one-shot PDs. Exploring the range of situations in which Simpson-type effects can occur in PD is an interesting direction for future work.

The experiment below contrasts a Mixed condition, where nice and nasty games are randomly sampled with equal frequency and two control conditions, one with *nice* games only and another with *nasty* games only. Each game was a one-shot prisoner's dilemma because agents were paired afresh for each game, over a computer network; and agents were anonymous. According to previous experimental results on game-playing (Rapoport & Chammah, 1965) it is natural to expect that cooperation will be highest in Nice condition, lowest in Nasty condition, and intermediate in Mixed condition. But according to the current analysis, when *nice* and *nasty* situations are mixed, there is an induced correlation between cooperativeness of self and other, which should boost cooperativeness. Thus, we predict that Mixed condition should produce as much, or more, cooperation than Nice condition.

Method

Participants. Twenty participants took part in each of the three conditions of this study (so there were 60 participants in total) recruited from the University of Oxford student population via the experimental economics research group mailing list. Participants were typically in the age-range 18-30. (All participants subsequently played a further 48 games, to explore a context effect not considered in this paper.) All participants were paid £3 plus performance related winnings of up to £3.

Design. As outlined above, there were three conditions in the experiment: Nice games only; Nasty games only; and a Mixed condition consisting of equal numbers of *nice* and *nasty* games combined randomly. In order to control for the effects related to the absolute magnitude of the received pay-off from each round, we used four different versions of these games: the pay-offs in Figure 3a and 3b were either used directly, or multiplied also by 4, 7, and 10; different pay-off values were randomly intermixed. In each condition, different participants played a sequence of 48 randomly chosen games.

Procedure. The experiment was conducted interactively in groups containing 4 to 8 participants. The participants were informed that on each round of the game they would play against a randomly selected player from their group. This random matching aimed to make it impossible to infer the strategy of the other player from the history of the game, and thus to create the one-shot situation of playing against a stranger. Thus we also aimed to prevent people from learning a model of their opponent, which is another significant contextual factor that has been shown to affect strategic behavior (see Pruitt and Kimmel, 1977, for a review).

Each condition consisted of a sequence of rounds of PD game in which players make their choices simultaneously. We used the label "1" to denote the cooperative response and "2" for the uncooperative one in order not to prime certain social values in the group, which might induce certain strategies that could additionally bias the results. The games were presented in a random order. On each round of the game the participants were presented with a matrix of the game on the computer screen and they had to choose their decision strategy (1 or 2), and after both players in each pair have made their decisions the round ends and they were informed on the screen about the decision made by the other player, and their pay-offs from the game. At the end of the experiment, the accumulated score in points was transferred into cash according to an exchange rate. In this experiment the participants were also asked to state how probable they think it is that the other player will cooperate, but we do not report these results here because our simulation was designed to model only the choice behaviour and also because the prediction results showed the same trend as the choice data.

To focus participants' attention on the differences between the games, it was explicitly stated in the instructions that in every round the pay-off values in the matrix will change, and that we are interested in how these changes influence people's decision strategy. There was also a detailed explanation of the strategic pay-off structure of the games and the differences between the two types of game, which we are here calling the *nice* and *nasty* games (although without, of course, using these evaluative labels). This aimed to eliminate an effect that we observed during pilot tests: that participants sometimes stop attending to each particular game, and start to play according to some (usually social or value-based) rule (e.g., start always to defect because they have learned in economics textbooks that this is the correct strategy).

Results

In order to see how each variable changed over time, we averaged the raw results per each participant in blocks of six rounds to reduce the variability of the data. So there were eight blocks for the 48 rounds. The figures presented here were derived by first calculating the results individually for each participant and then averaging over all participants per condition.

Figure 4 presents the average cooperation rate in each condition. The average cooperation in the Mixed condition (.45) was higher than both the Nice condition (.33) and the Nasty condition (.18). The differences between the conditions were examined with a one-way univariate analysis of variance, with the mean cooperation as the dependent measure. As predicted by our account, cooperativeness follows the ordering Mixed > Nice > Nasty, and there was a significant main effect of condition, F(2, 57) = 9.27, p = .0003.

INSERT FIGURE 4 ABOUT HERE

Our explanation for this effect requires that agents' choices are correlated in the Mixed condition. This is confirmed in Table 1, which shows that there was a strong tendency to cooperate in Nice games (more cooperation than defection); and not to cooperate in Nasty games (more defection than cooperation). Hence, in the Mixed condition, the exogenous variable of 'game-type' induces a correlation between the agents' choices. By contrast (as can be read off Figure 4), in the conditions where agents had only Nice or Nasty games, there was more defection than cooperation.

INSERT TABLE 1 ABOUT HERE

Table 2 shows that, as expected, in the Mixed condition only, the average pay-off on trials where an agent chose to cooperate was much higher than the average pay-off on trials where the agent chose not to cooperate. Thus, the results presented in Table 2 show that the correlation between responses in the Mixed condition caused the average received pay-off to be higher on trials on which an agent chooses *C* than when the agent chooses *D*. This does not occur in the Nice or Nasty conditions.

INSERT TABLE 2 ABOUT HERE

Table 3 shows the average pay-offs for *C* and *D* in Mixed condition broken down by Nice and Nasty games. Note that the pay-off for *C* is lower than *D* in both Nasty and Nice games, but the overall pay-off for *C* is higher when all data are summarized together (i.e., the weighted average of the pay-off received from the two games) because the relative proportion of *C* play is higher than *D* play; which is a clear example of Simpson's paradox.

INSERT TABLE 3 ABOUT HERE

The Simpson's paradox arising in the Mixed condition is best summarised and graphically illustrated in Figure 5. Nice games and Nasty games represent the average pay-off received from the nice and nasty games when a player plays *C* or *D* respectively. However, the average pay-off received from *C* play in Nice games is higher than the pay-off received from *D* play in Nasty games because when players play Nice games they tend to play *CC* most of the time, and while when they play Nasty games they end up playing *DD* most of the time. As a consequence of these skewed distributions, the average pay-off reinforcement for the *C* strategy is higher than for the *D* strategy. In summary, the Simpson's paradox has biased each player's judgment of the utility of each strategy by enforcing the conclusion that overall *C* was more profitable than *D* although in each particular game *D* was the strategy that brought higher pay-off.

INSERT FIGURE 5 ABOUT HERE

We believe that average pay-off (or reinforcement) plays a vital role in sustaining the stable cooperation rate.

Discussion

People evaluate actions on the basis of their average pay-offs (e.g., Erev & Roth, 1998, 2002, apply such reinforcement learning models to behaviour in games), but they have a cognitive blindspot: they cannot without correcting for biased sampling. When situations differ in the degree to which they encourage cooperation, (un)cooperative actions tend to be associated with (un)cooperative actions of others. These associations may entrench altruistic cooperative behaviour with strangers, even where selfish behaviour has a higher pay-off in every situation; and, moreover, may lead to an illusory belief that kindness to strangers begets other strangers' kindness to the self. Thus, a cognitive blind-spot, our inability to correct for sampling bias, may be an important foundation for social behaviour.

Multi-Agent Simulation

The empirical data reviewed above indicate that people evaluate actions on the basis of their average pay-offs, and are unable to correct for biased sampling (Fiedler, 2000). These results are also consistent with the idea that people learn to choose their responses according to the reinforcement for each choice option. Indeed, Erev and Roth have applied such reinforcement learning models to behaviour in games (Erev & Roth, 1998; Erev & Roth, 2002), showing good fits with a range of empirical data on games including repeated PD games. Reinforcement learning has not been applied to sequences of one-shot PD however---indeed, except in the light of the Simpson's paradox observations that we have made here, it might seem pointless to do so, because it might be assumed that players would inevitably converge to play *D*. The goal of the present simulations is to investigate whether a population of reinforcement learners can model the experimental results that we have observed with human players who stably choose high levels of *C* responding.

To try to describe the experimental results presented in the previous section, we propose a probabilistic choice model that uses reinforcement learning to updates the choice probabilities for *C* and *D* respectively. The model is determined a set of recursive relations, between choices and pay-offs. The pay-offs of the games determine how agent's choices deteremine the pay-offs they received. But more subtle is the relationship between the "reinforcement" history of pay-offs determine the current choices. Let us consider this process in more detail.

Let us start by capturing the overall amount of cooperation for *C* and *D*, based on the previous pay-offs for those choices. The simplest assumption is simply to adopt the most elementary form of Herrnstein's matching law (see Rachlin & Laibson, 1997, for a review), and assume that:

$$\Pr(C) = \frac{\hat{M}(C)}{\hat{M}(C) + \hat{M}(D)}$$

$$\Pr(D) = \frac{\hat{M}(D)}{\hat{M}(C) + \hat{M}(D)} \qquad (1)$$

where Pr(*C*) is overall probability of playing *C*, independent of whether the agent is playing a nice or a nasty game. The $\hat{M}(C)$ and $\hat{M}(D)$ is the average pay-off for playing *C* and *D* respectively. We now need to embody in the model the key empirical observation by Rapoport and Chammah (1965) that 'nice' games systematically lead to more cooperation than 'nasty' games. Specifically, we assume that the cooperation level in nasty games is a fraction *k* of the cooperation level in nice games, where 0<*k*<1.

$$\Pr(C \mid nasty) = k \Pr(C \mid nice) \tag{2}$$

We can now solve for Pr(*C*|*nasty*) and Pr(*C*|*nice*) given the further constraint, which is simply an identity of probability theory, that the probability of playing *C* is a weighted sum of Pr(*C*|*nasty*) and Pr(*C*|*nice*), weighted by the probabilities of each type of game, and similarly for Pr(*D*):

$$P(C) = P(C \mid nice)P(nice) + P(C \mid nasty)P(nasty)$$

$$P(D) = P(D \mid nice)P(nice) + P(D \mid nasty)P(nasty) \tag{3}$$

Intuitively, in this model, the matching law determines the overall level of cooperativeness. But in addition, there is further bias to play *C* is nice games, enforced by (2) and (3). Rearranging, we can solve for the levels of cooperation in each type of game, given quantities that can be directly observed in the simulation (the averages $\hat{M}(C)$, $\hat{M}(D)$, the probabilities of each type of game, Pr(*nice*), Pr(*nasty*), and *k*).

$$P(C \mid nice) = \frac{\hat{M}(C)}{\hat{M}(C) + \hat{M}(D)} . \frac{1}{P(nice) + kP(nasty)}$$

$$P(C \mid nasty) = \frac{\hat{M}(C)}{\hat{M}(C) + \hat{M}(D)} . \frac{k}{P(nice) + kP(nasty)} \tag{4}$$

Finally, we note that there are some minor additional points to be noted in relation to *k*. If we fix *k*, then this constrains the overall level of cooperation in nasty games to be at most *k* (this value will be reached, of course, when cooperation in nice games reaches 1). We chose to avoid this problem by adaptively modifying *k*, depending on the overall cooperation level of the game.

$$k = k_0 + (1 - k_0)\max\left(0, \frac{P(C) - P(nice)}{1 - P(nice)}\right) \qquad (5)$$

where k_0 is a fixed parameter between 0 and 1. The key observation here is that as Pr(*C*) increases, *k* also increases, and hence the difference between the cooperation levels in the *nasty* and *nice* games can become arbitrarily small. The Appendix provides a brief rationale for the specific implementation of this approach used here.

Overall, then, the equations in (4) show how to determine the cooperation levels in the nice and nasty games; these cooperation differ by a factor of *k*, which is modified on-line in the simulations by (5).

These equations determine the set of rules needed to play the game. The quantities in equation (4) are determined by the actual pay-offs during the simulation, *U*(*C*) and *U*(*D*), and the preset distribution of *nice* and *nasty* games, P(*nice*) and P(*nasty*). In order to start the procedure we only need to specify initial values for *P*(*C* |*nice*) and k_0 (this starting value of *k*). We specified the initial values for *P*(*C* |*nice*) = .5 and a value for k_0 = .1 (this figure seemed reasonable, given the heuristic justification that *nasty* games were ten times less cooperative than the *nice* games according the cooperation index proposed by Rapoport and Chammah, 1965). Note that the ratio P(C|*nasty*)/P(C|*nice*) is fixed to 0.1 only in the beginning of the simulation and is changing with increasing P(C|*nice*). In the simulation presented here, *k* was observed to asymptote at about 0.2-0.3, which is consistent with this parameter estimated from the experimental results. The precise starting value of the bias, k_0, was not crucial. The qualitative pattern results reported below, which occurs nearly always with the values we chose here, arises less than 50% of the time, if the initial value of k_0 exceeds .3.

There were three conditions, Mixed, Nice, and Nasty, as in the experiment. The averaging was performed over the previous ten games played in a sequence. The simulations were performed using the multi-agent platform WSIM, under development at the New Bulgarian University. We tried to be as close as possible to the experiment and used the same games and set up. The games were not multiplied by 4, 7, and 10, as in the experiment, because the absolute magnitudes of the pay-offs do not make any difference for the learning and choice algorithm described here. However, in the simulation all games were rescaled to have positive pay-offs, because the model calculates absolute reinforcements and does not differentiate between positive and negative values. To mimic the experiment, eight artificial agents per group played PD game against different opponent on each round. At the beginning of the session, each agent played with initial probability of cooperation, P(C|*nice*) and/or P(C|*nasty*) depending on the condition, and updated this probability depending on the received reinforcement (pay-offs), as formulated in Equation (4).

Results

Figure 6 presents the average cooperation rate in each condition of the simulation. We averaged the raw results per each agent in blocks of six rounds so there were eight blocks for the 48 rounds (as we did with the experimental results). The figures were derived by first calculating the results for each agent and then averaging over all participants per condition. The average cooperation in the Mixed condition (.44) was higher than both the Nice condition (.34) and the Nasty condition (.25). Thus, the expected effect observed in the experiment, namely higher cooperation in the Mixed condition, was also found in the simulations. The differences between the conditions were examined with a one-way univariate analysis of variance, with the mean cooperation as the dependent measure. There was a significant main effect of condition, F(2, 57) = 18.3, p < .0001.

INSERT FIGURE 6 ABOUT HERE

Our hypothesis about the explanation of the increase of cooperation in the Mixed condition implies coordination between the agents due to the bias to cooperate more in *nice* games. This coordination should lead to a relatively high percentage of *C* play in the *nice* games and high percentage of *D* play in the *nasty* games. The results presented in Table 4 confirm this hypothesis. There was a strong tendency to cooperate more in Nice games and not to cooperate in Nasty games. This result is very similar to the experimental result shown in Table 1. Thus the endogenous bias to be more cooperative in *nice* games allowed the exogenous variable of 'game-type' to induce a correlation between the agents' choices.

INSERT TABLE 4 ABOUT HERE

However, can this initial bias alone explain an average cooperation rate in the Mixed condition that is higher than that observed in the Nice condition? It cannot, because only in half of the games in the Mixed condition, the cooperation rate will be as high as in the Nice condition, while in the other half the cooperation will be much lower. As a result the average cooperation in the Mixed condition would be lower than in the Nice condition. However, note that in our model choice behaviour is updated by reinforcement learning that takes into account only the average received pay-offs per strategy without taking into account the difference between the games. This algorithm allows, if choices are coordinated, to receive higher average pay-off for playing *C* than for playing *D* in the Mixed condition and vice versa in the Nice and Nasty conditions. This is because when an agent plays *C* the outcome is usually *CC*, while when *D* is played the outcome is usually *DD* (and *CC* get bigger pay-off than *DD* according to the structure of the PD game). Table 5 summarises this result in a format that is comparable with Table 2 from the experiment above. Thus this result is consistent with the findings of the experiments. In summary, our analysis shows that the result in the Mixed condition is caused by the higher pay-offs received from *C* play in comparison to *D* play (similarly to the experiment), which does not occur in the Nice or Nasty conditions.

INSERT TABLE 5 ABOUT HERE

Table 6 shows the average pay-offs for *C* and *D* in Mixed condition separately for Nice and Nasty games. The pay-off for *C* is lower than *D* in both Nasty and Nice games, but the overall pay-off for *C* is higher when all pay-offs are averaged together. This because the relative proportion of *C* play is higher than *D* play; which is a clear example of Simpson's paradox. Thus this result replicates the experimental data.

INSERT TABLE 6 ABOUT HERE

The modelling results clearly demonstrate how reinforcement learning can sustain and increase cooperation due to its inability to bring together samples of different origin and different levels of analysis that can support different conclusions and strategies (see Fiedler, 2000, for a discussion of other judgment biases arising from such biased sampling). This effect is known as the Simpson's paradox and, as we claimed before, any learning algorithm that is based on average reinforcement, cannot resolve this sampling problem.

Overall, both the simulation results and the experimental data indicate that inducing a correlation between agents' behaviour, in this case by varying the cooperativeness and the pay-offs of the games in the Mixed condition, can lead to gradually increasing cooperation. Moreover, the cooperation rate where there is a mix of games can be consistently higher than the cooperation rate when there are only *nice* games. That is, by adding *nasty* games to the mix of games that agents encounter, where *nasty* games are those that, considered alone, tend to produce uncooperative response, can actually increase the overall amount of cooperation. This apparently paradoxical effect results directly from the operation of Simpson's paradox. To the degree that artificial, animal or human agents are influenced in their choices by reinforcement, then they will fall into Simpson's paradox because reinforcement learning algorithms cannot 'understand' simultaneous sampling at different levels (which is necessary to solve the paradox). So, if the responses of players are correlated due to some common factor (here, whether the game itself is *nice* or *nasty*), they may potentially fail to find the Nash equilibrium, and instead stabilize on playing the cooperative (even though dominated) strategies.

General Discussion

We began by noting that the concept of Nash equilibrium from game theory, the normative economic framework for understanding how agents should interact, faces three types of challenge: too many Nash equilibria; observed behaviour that does not seem to fit any Nash equilibrium; and strong rationality assumptions, which may overestimate the computational power both of artificial agents, and the cognitive system.

We saw that one way of attempting to finesse the problem of cognitive limitation is to model agents as reinforcement learners---and in many circumstances, such agents will converge to Nash equilibria. But we noted also that some patterns of observed behaviour do not seem to converge in this way---and indeed, the human experiment that we discuss here is an example of this phenomenon. We believe that it is likely that there is a wide range of factors that are relevant to understanding why people do not follow Nash strategies, including effects of cultural and moral norms. But we have demonstrated in an experiment and also using computational modelling, that the basic principle of reinforcement learning provides a further, and perhaps rather unexpected, mechanism for stable non-Nash strategies like mutual cooperation in a one-shot PD game.

From the point of view of understanding the behaviour of populations of interacting animal and human agents, it is interesting to ask how likely the type of effect that we have considered might occur. Specifically, how often are there factors that lead to the choices of the players in a PD-type game being positively correlated (those opening up the conditions for Simpson's paradox to operate, as we have discussed). We suggest that these circumstances may be quite widespread. To the degree that agents are similar; and to the degree to which the factors that determine their decision are related, it seems likely that there will be likely to be a positive correlation (and perhaps quite a large positive correlation) between their behaviours.

We have already noted earlier that one source of such a correlation might be a tendency for agents in the same group (however 'group' may be defined) to cooperate---and effect that has been widely studied experimentally (Hewstone, Stroebe & Stephenson, 1998). But purely environmental factors might play the same role. If it turns out that agents cooperate over food distribution more or less effectively, depending on the nature of the food to the divided, or depending on the degree of food shortage that is prevalent, or for any other reason, then these reasons will bear on both agents equally. Hence, it will be likely that positive correlations between agents' choices will be observed.

In many real-world settings, of course, the strict framework of the one-shot PD will not hold. Thus, in small human or animal communities it is likely that agents may play the same game against each other repeatedly; and that, for some species at least, they will be able to track both the identity, and some elements of the past behaviour, of other agents. In such cases, interactions may be more appropriately modelled in terms of repeated PD, rather than the one-shot PD under study here. Nonetheless, it is still likely that the more sophisticated memory and reasoning processes required to support this kind of behaviour will not completely overwrite the effects of simple reinforcement learning (Erev & Roth, 2000). Hence, we suggest that the present results may be of relevance to understanding cooperative behaviour even in domains where interactions between agents are not strictly one-shot---essentially because reinforcement learning appears to be a powerful learning mechanism in the biological domain; and reinforcement learning treats each game as one-shot.

We hope that experimental work and computer simulations with computationally simple agents may integrate in important ways with the economic, game-theoretic style of explanation of how agents should interact. Returning to the three issues concerning the applicability of economic concepts such as the Nash equilibrium that we noted before, we believe that experimental and computational simulations of the processes of learning that lead to animals or people's behaviour may help understand (1) why some Nash equilibria are favoured over others; (2) how, as in the simulations presented here, non-Nash equilibria can be reached; (3) under what conditions a purely rational explanation, which appears to impute to agents very substantial reasoning abilities, can, or cannot, be assumed to fit with the empirical results obtained with computationally limited agents.

We have, in particular, found a new mechanism which may contribute to explaining the stability of cooperation with strangers, who will never be encountered again, which seems to be a fundamental aspect of human social behaviour.

References

Axelrod, R. (1984). *The Evolution of Cooperation.* Basic Books, New York.

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

Bacharach, M. (1999). Interactive team reasoning: A contribution to the theory of co-operation. *Research in Economics, 53,* 117-147.

Baron, J. (1997). The illusion of morality as self-interest: A reason to cooperate in social dilemmas. *Psychological Science, 8,* 330-335.

Binmore, K. (1994). *Game theory and the social contract. Volume 1: Playing fair*. Cambridge, MA: MIT Press.

Colman, A. (2003). Cooperation, psychological game theory, and limitations of rationality in social interaction. *Behavioral and Brain Sciences, 26,* 139-153.

Cooper, R., DeJong, D. V., Forsythe, R. & Ross, T. W. (1996). Cooperation without Reputation: Experimental Evidence from Prisoner's Dilemma Games. *Games and Economic Behaviour, 12,* 187-218.

Curley, S. P. & Browne, G. J. (2001). Normative and descriptive analysis of Simpson's Paradox in decision making. *Organizational Behavior and Human Decision Processes, 84,* 308-333.

Dawes, R. & Thaler, R. (1988). Cooperation. *Journal of Economic Perspectives, 2,* 187-197

Dawkins, R. (1976). *The Selfish Gene*. Oxford University Press, Oxford.

Erev, I. & Roth, A.E. (1998). Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria, *American Economic Review, 88,* 848-881.

Erev, I. & Roth, A.E. (2002). Simple Reinforcement Learning Models and Reciprocation in the Prisoner's Dilemma Game. In *Bounded Rationality: The Adaptive Toolbox* (eds Gigerenzer, G., & Selten, R.) (pp. 215-231). Cambridge, MA; London: MIT Press.

Fiedler, K. (2000). Beware of samples! A cognitive-ecological sampling approach to judgment biases. *Psychological Review, 107,* 659-676.

Fiedler, K., Walther, E., Freytag, P. & Nickel, S. (2003). Inductive reasoning and judgment interference: Experiments on Simpson's paradox. *Personality and Social Psychology Bulletin, 29,* 14-27.

Fiedler, K., & Juslin, P. (Eds) (2006). *In the beginning there is a sample: Information sampling as a key to understand adaptive cognition*. Cambridge, UK: Cambridge University Press.

Fudenberg, D. & Tirole, J. (1996). *Game theory*. Cambridge, MA: MIT Press.

Gold, N. (Ed.) (in press). *Teamwork*. London: Palgrave.

Hamilton, W. D. (1964). The genetical evolution of social behaviour, I and II. *Journal of Theoretical Biology, 7,* 1-52.

Hewstone, M. Stroebe, W. & Stephenson, G. M. (Eds.) (1996). *Introduction to social psychology* (2nd edition). Oxford: Blackwell.

Hurley, S. (1989). *Natural reasons*. New York: Oxford University Press.

Hurley, S., & Chater, N. (Eds.) (2005). *Perspectives on Imitation*: *From neuroscience to social science* (2 volumes). Cambridge, MA: MIT Press.

Kreps, D. (1990). *A course in microeconomic theory*. Princeton, NJ: Princeton University Press.

Nowak, M. A. & May, R. M. (1992). Evolutionary games and spatial chaos. *Nature, 359,* 826-829.

Nowak, M. A. & Sigmund, K. (1989). Oscillations in the evolution of reciprocity. *Journal of Theoretical Biology, 137,* 21-26.

Pearce, J. M. & Bouton, M. E. (2000). Theories of associative learning in animals. *Annual Review of Psychology, 52,* 111-139.

Rachlin, H., & Laibson, D. I. (1997). *The Matching Law: Papers in Psychology and Economics by Richard J. Herrnstein*. Cambridge, MA: Harvard University Press.

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

Satow, K. L. (1975). Social approval and helping. *Journal of Experimental Social Psychology, 11,* 501-509.

Schneider, F. W. & Mockus, Z. (1974). Failure to find a rural-urban difference in incidence of altruistic behavior. *Psychological Reports, 35,* 294.

Simpson, E. H. (1951). The interpretation of interaction and contingency tables. *Journal of the Royal Statistical Society, 13,* 238-241.

Sober, E. & Wilson, D. S. (1998). *Unto others: The evolution and psychology of unselfish behaviour*. Cambridge, MA, US: Harvard University Press.

Spellman, B. A. (1996). Acting as intuitive scientists: Contingency judgments are made while controlling for alternative potential causes. *Psychological Science, 7,* 337-342.

Spellman, B. A., Price, C. M., & Logan, J. M. (2001). How two causes are different from one: The use of (un)conditional information in Simpson's paradox. *Memory and Cognition, 29,* 193-208.

Sutton, R.S., Barto, A.G. (1998). *Reinforcement Learning: An Introduction*. MIT Press.

Thorndike, E. L. (1911). Animal intelligence: An experimental study of the associative processes in animals. *Psychological Review Monograph Supplement*, 2: Whole No. 8.

Waldmann, M. R. (1996). Knowledge-based causal induction. In *Causal learning* (eds Shanks, D. R., Holyoak, K., & Medin, D. L.) 47-88. San Diego, CA, US: Academic Press.

Wedekind, C., & Braithwaite, V. A. (2002). The long term-benefits of human generosity in indirect reciprocity. *Current Biology 12,* 1012-1015.

Table 1

Percentage of C and D responses in the Mixed condition

Decision	Nice	Nasty
Cooperate (C)	37%	8%
Not-Cooperate (D)	13%	42%

Table 2

Average pay-offs for C and D in the Mixed, Nice and Nasty conditions

Decision	Nasty	Nice	Mixed
Cooperate (C)	-42.7	20.1	28.7
Not-cooperate (D)	-31.3	21.1	-15.8

Table 3

Average pay-offs for C and D in the Nice and Nasty games in the Mixed condition

Decision	Nasty	Nice
Cooperate (C)	-35.4	41.2
Not-cooperate (D)	-30.5	46.4

Table 4

Percentage of C and D responses in the Mixed condition of the simulation

Decision	Nice	Nasty
Cooperate (C)	37%	7%
Not-Cooperate (D)	8%	48%

Table 5

Average pay-offs for C and D in the Mixed, Nice and Nasty conditions of the simulation

Decision	Nasty	Nice	Mixed
Cooperate (C)	0.88	1.28	3.16
Not-cooperate (D)	2.92	2.17	2.88

Table 6

Average pay-offs for C and D in the Nice and Nasty games in the Mixed condition of the simulation

Decision	Nasty	Nice
Cooperate (C)	0.72	3.62
Not-cooperate (D)	2.60	4.39

Figure Captions

Figure 1. Prisoner's Dilemma Game matrices (basic versions).

Abstract version of the game. Here, the left of the pair of items is the pay-off for agent 1; the right hand item is the pay-off for agent 2. The game is defined by the chain of inequalities Tempt >Co-op >Defect >Sucker, where Co-op is the pay-off if both cooperate (C), Defect is the pay-off if both defect (D), Tempt is the pay-off if one defects and the other cooperates (i.e., it represents the "temptation" of defecting against a co-operative opponent), and Sucker is the pay-off if one cooperates and the other defects. From an individualistic perspective, defection always pays (because Tempt > Co-op; and Defect > Sucker). But, in contrast, it is better if both agents cooperate (with a Co-op, Co-op outcome) than if both agents defect (with a Defect, Defect outcome), because Co-op > Defect.

Figure 2. Simpson's paradox.

Although the overall contingency between treatment (inhaler vs. no inhaler) and breathing improvement is negative in the population (upper table), the contingency is positive within both subsets (people with asthma and people without asthma). In order to solve this paradox people have to understand that they are simultaneously sampling at different levels.

Figure 3. Prisoner's Dilemma Game matrices (basic versions).

a) Positive/Nice version of the game. b) Negative/Nasty version of the game. As in Figure 1, the left hand side number of a pair of outcomes is that associated with "you"; the right hand number is the outcome to the "other."

Figure 4. Cooperation rate during the 48 trials in the Nice, Nasty, and Mixed conditions.

Figure 5. Average pay-off received from Nice and Nasty games when a player chooses C and D respectively.

Figure 6. Cooperation rate during the 48 trials in the Nice, Nasty, and Mixed conditions of the simulation.

Figure 1

		Agent 2	
		C	*D*
Agent 1	*C*	Co-op, Co-op	Sucker, Tempt
	D	Tempt, Sucker	Defect, Defect

Figure 2

	Predictor	Dependent Variable		Conditional	Contingency
		Clear airway	Blocked airway	P(Clear)	
Overall	Inhaler	499	501	.50	-.32
	No inhaler	820	180	.82	
People with Asthma	Inhaler	400	500	.44	.34
	No inhaler	10	90	.10	
People Without Asthma	Inhaler	99	1	.99	.09
	No inhaler	810	90	.90	

Figure 3

a) Nice game

		Other	
		1	*2*
You	*1*	10, 10	0, 11
	2	11, 0	1, 1

b) Nasty game

		Other	
		1	*2*
You	*1*	-5, -5	-11, 0
	2	0, -11	-6, -6

Figure 4

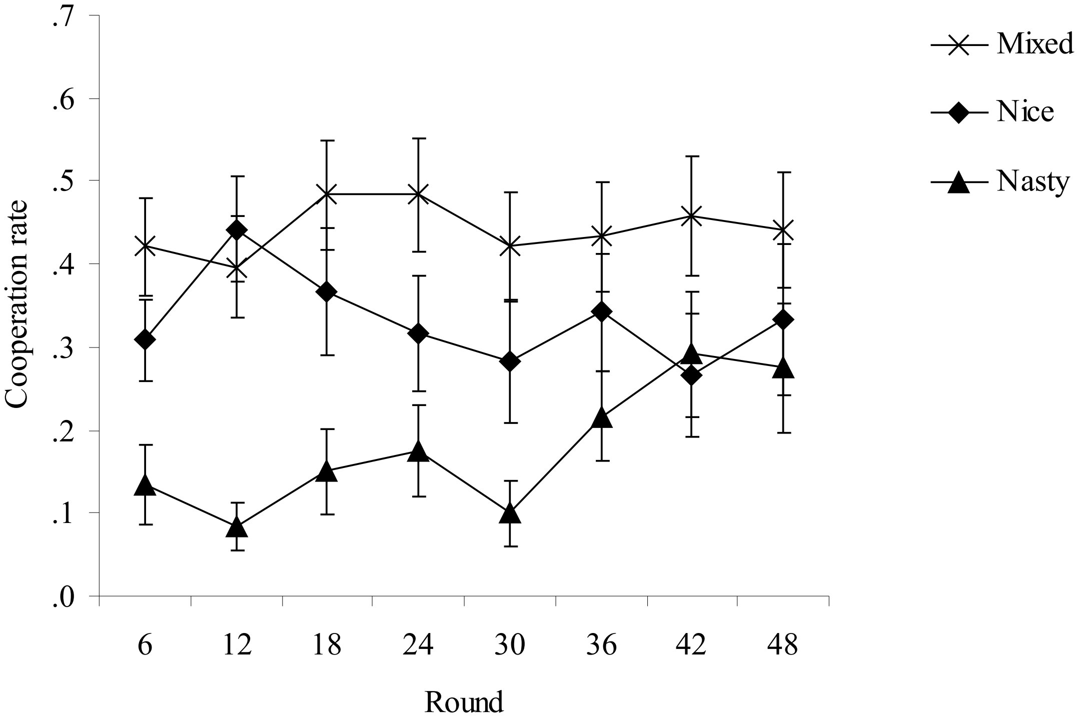

.7
.6
.5
.4
.3
.2
.1
.0
Cooperation rate
6
12
18
24
30
36
42
48
Round
Mixed
Nice
Nasty

Figure 5

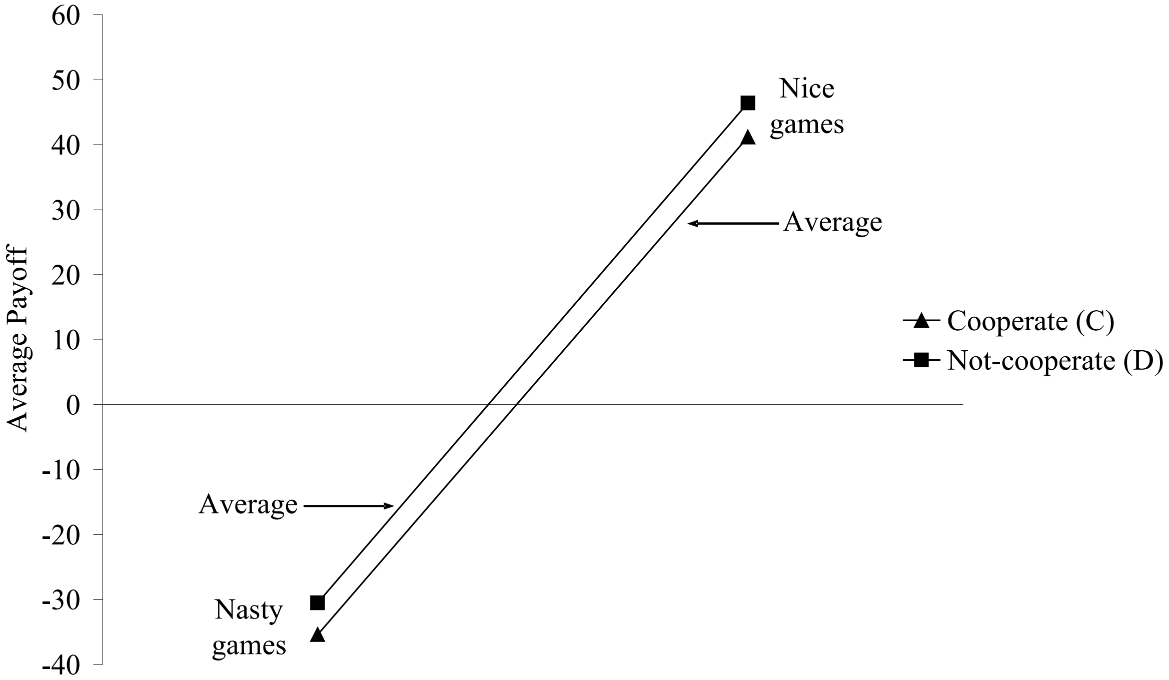
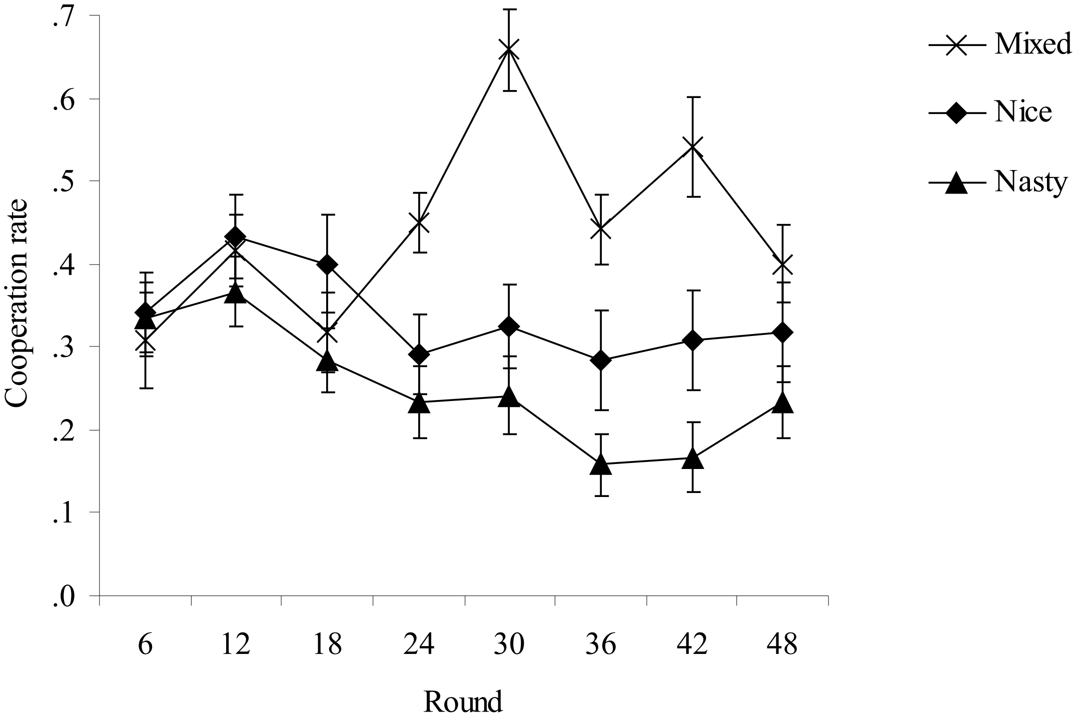

60
50
40
30
20
10
0
-10
-20
-30
-40
Average Payoff
Nice games
Average
Average
Nasty games
▲ Cooperate (C)
■ Not-cooperate (D)

Figure 6



.7
.6
.5
.4
.3
.2
.1
.0
Cooperation rate
6
12
18
24
30
36
42
48
Round
Mixed
Nice
Nasty

Appendix

If Pr(*C*|*nasty*) is a proportion *k* of Pr(*C*|*nice*), then if *k* is fixed, this implies that Pr(*C*|*nasty*) cannot exceed *k*, and will attain this value when Pr(*C*|*nice*) = 1. This, moreover, will constrain the overall cooperation rate Pr(*C*) to be no greater than Pr(*nice*) + *k*Pr(*nasty*) (using equation (3)). This can be rewritten as:

$$P(C) \leq \Pr(nice) + k(1 - \Pr(nice) \qquad \text{(A1)}$$

and hence

$$k \geq \frac{\Pr(C) - \Pr(nice)}{1 - \Pr(nice)} \qquad \text{(A2)}$$

where we have the additional constraints that *k* is less than or equal to 1 (this captures the fact that cooperation rates are assumed to be higher in nice games); and of course, *k* must be greater than or equal to 0, or Pr(*C*|*nasty*) will become a negative probability. Note that Pr(*C*) - Pr(*nice*) can be negative so that to express these constraints on *k* we need to write:

$$1 \geq k \geq \max\left(0, \frac{\Pr(C) - \Pr(nice)}{1 - \Pr(nice)}\right) \qquad \text{(A3)}$$

Thus, to avoid *k* becoming too low to fit with Pr(*C*), we can reparameterize *k* as running between these maximum and minimum values:

$$k = k_0 + (1 - k_0)\max\left(0, \frac{\Pr(C) - \Pr(nice)}{1 - \Pr(nice)}\right) \qquad \text{(A4)}$$

Where k_0 is a parameter between 0 and 1, which determines the size of the difference between cooperativeness in nice and nasty games. If k_0 = 1 the level of cooperativeness is the same in both nice and nasty games; as k_0 approaches 0, then difference between cooperation rate increases. Nonetheless, with this parameterization, the overall cooperation rate Pr(C) can vary between 0 and 1.