politics, philosophy & economics article

London
Thousand Oaks, CA
and New Delhi

1470-594X
200802 7(1) 81–98

Explaining fairness in complex environments

Kevin J.S. Zollman
University of California, Irvine, USA

abstract This article presents the evolutionary dynamics of three games: the Nash bargaining game, the ultimatum game, and a hybrid of the two. One might expect that the probability that some behavior evolves in an environment with two games would be near the probability that the same behavior evolves in either game alone. This is not the case for the ultimatum and Nash bargaining games. Fair behavior is *more* likely to evolve in a combined game than in either game taken individually. This result confirms a conjecture that the complexity of our actual environment provides an explanation for the evolution of fair behavior.

keywords evolutionary game theory, Nash bargaining game, ultimatum game, fairness

1. Introduction

In the game theoretic study of justice, two games have predominated in the literature: the Nash bargaining game and the ultimatum game. In the Nash bargaining game two players are given a windfall. Each player makes a demand, and if the two demands do not exceed the total good, both receive their demand. Otherwise, both receive nothing. In the ultimatum game, two players again receive a windfall, but only one player suggests a division. After learning of the first player's proposal, the second must either accept or reject it. If the second accepts, both receive the amounts suggested by the first, otherwise they receive nothing. Both games provide an interesting context in which to study fair behavior, a central feature of our conceptions of justice.[1] Traditionally, each of these games is modeled in isolation, with the hope that plausible assumptions about the structure of the populations will result in the evolution of fair play.[2] While this modeling strategy has resulted in several evolutionary models which predict fair

DOI: 10.1177/1470594X07081299
Kevin J.S. Zollman is a graduate student in the Department of Logic and Philosophy of Science at the University of California, Irvine, USA [email: kzollman@uci.edu]

behavior in the Nash bargaining game, there are fewer successful models of the ultimatum game.[3]

It is uncontroversial that humans confront situations similar to both games (and perhaps many more) in regular strategic interactions and that norms of fairness govern behavior in both. By studying these games in isolation, scholars implicitly hope that the explanation of fair behavior for each game individually will carry over into the more complex environment that humans actually confront. In fact, in response to the failure to explain adequately the emergence of cooperation in the ultimatum game, some scholars have suggested that the presence of other games (such as the Nash bargaining game) have helped cooperation along. In order for this to be possible, it must be the case that the probability cooperation evolves in a complex situation (one that involves both games) would be more like the Nash bargaining game than the ultimatum game. A priori, one might expect that the result for a complex strategic interaction will lie somewhere in between the simple models taken individually, that is, the actual complexity will not alter the strategic situation radically. This is not the case, however. In fact, for many combinations the probability that fair behavior evolves is *greater* than the probability that fair behavior will evolve in either of the games taken independently.

In Section 2, the evolutionary dynamics of the Nash bargaining and ultimatum games will be analyzed. Here we find that the evolution of fair behavior in the Nash bargaining game is relatively likely. It is rather more difficult to model the evolution of fair behavior in the ultimatum game. A suggestion for modeling fair behavior in the ultimatum game as a norm of fairness will be discussed in Section 3. Section 4 offers a formal model where players regularly confront both the ultimatum and Nash bargaining games. The evolutionary dynamics of this game are analyzed and an anomalous result is explained. Finally, Section 5 concludes.

2. Two games

2.1. The Nash bargaining game

In the Nash bargaining game each player must choose a number (a demand) between zero and the total good. If the sum of the two demands is not greater than the total good, each player receives her demand, otherwise both players receive nothing. There are many Nash equilibria in the Nash bargaining game. Any two proposals which sum to the total good are a Nash equilibrium. Neither player would benefit by increasing her demand, since that would result in her receiving nothing, nor would she benefit from reducing her demand, since that would result in her receiving less than she might have otherwise.[4] Since there are many Nash equilibria in this game, we are confronted with a problem: how do individuals select an equilibrium? Unsurprisingly, when confronted with this circumstance most people choose the 50–50 split,[5] but what about that equilibrium makes it salient?

One possibility for explaining the near universality of this solution is to study the evolutionary dynamics of populations of players. As an example, suppose we restrict proposals to ⅓, ½, and ⅔. With this restriction there are two evolutionary stable states (ESSs).[6] In one the entire population demands ½. In this population, the average payoff is ½; since a mutant who proposes ⅓ or ⅔ would receive less, neither can invade the population. In the other ESS, half of the population demands ⅓, the other ⅔ (a polymorphism). Here, the average payoff to the population is ⅓; the ⅓ proposers always receive what they propose and the ⅔ proposers receive their demand half of the time (the other half they meet themselves and receive nothing). Any increase in the number of ⅔ proposers would reduce the payoff to ⅔ proposers, driving the population back to the polymorphic equilibrium. Similarly, any increase in the number of ⅓ proposers would result in ⅔ proposers meeting compatible types more often and thus receiving a higher payoff, again driving the population back to the polymorphic equilibrium. Finally, a ½ proposer cannot invade since she will only meet a compatible type approximately half of the time and thus receive an average payoff of ¼. There is an equilibrium where half the population proposes ⅓, one-sixth proposes ½, and the remaining one-third proposes ⅔. This population is not evolutionarily stable because any increase in the proportion of any of the strategies will drive the population away from that state.

Brian Skyrms uses the standard replicator dynamics in which an individual strategy grows in proportion to its payoff in the population.[7] Sampling at random from the initial population proportions, he finds that the percentage of populations that converge toward fair proposals is not small: approximately 62 percent of the initial starting populations. The remaining populations are driven to the polymorphic equilibrium. Adding further assumptions about the structure of the population can increase the basin of attraction of fair behavior even more.[8]

On the basis of this game alone, it seems that we have a relatively good explanation of fair behavior. However, since it seems plausible that humans occasionally confront bargaining situations that are more like the ultimatum game than the Nash bargaining game, we might also want to investigate the properties of that game as well.

2.2. The ultimatum game

In the ultimatum game, one player must choose a number between zero and the total good, while the other must chose which demands to accept or reject. Like the Nash bargaining game, the ultimatum game has several Nash equilibria. Any strategy set where the first player proposes to keep the most that the second will allow is a Nash equilibrium. The first proposer would not want to propose less, since this would result in her receiving less, nor would she want to propose more because then she would receive nothing (since, by hypothesis, the second will refuse any split that gives the first more). The second player will do no better by rejecting the offer, since that would result in her receiving less.[9] Despite this large

number of Nash equilibria, only one set of strategies collectively satisfies a further restriction on players' rationality: sequential rationality.[10] A strategy is sequentially rational if at no point a player takes an action that guarantees she will receive a lower payoff than she would receive by taking another available action. If the proposer offers a split which gives the second any positive amount, the second does strictly worse by refusing the offer. So, no rejection strategies are sequentially rational. Knowing this, the first player ought to offer the smallest amount possible to the second player.

It is well known that despite this relatively simple reasoning, in experiments, players do not play sequentially rational strategies. In fact, players usually offer more than the smallest possible offer and low offers are occasionally rejected. Oosterbeek, Sloof, and Van de Kuilen perform an analysis on most of the available datasets and find that the average offer to the second player was 40 percent of the good and 16 percent of offers were rejected.[11] In an extensive cross-cultural study of small cultures ranging from the Amazon basin to the Indonesian archipelago, Henrich et al. observed a wide variety of strategies employed in the ultimatum game.[12] While some cultures did more closely approximate sequentially rational play, most did not.

These results indicate something very important. Behavior in the ultimatum game is culturally contingent and usually not sequentially rational. The important challenge raised by Henrich et al. is to construct a model that allows sequentially irrational play to evolve, but ensures that its evolution depends on particular features of the environment in which the norm evolves.

Our task is then to model the evolution of fair behavior in the ultimatum game. We might think of strategies in the ultimatum game as prior commitments which need not be sequentially rational. Perhaps an agent adopts a range of values that she considers reasonable and then accepts only those proposals that are within that range. If we restrict the game to three demands (⅓, ½, and ⅔) and three ranges of acceptability ([⅓, 1], [½, 1], and [⅔, 1]), we transform the two-stage ultimatum game into a simultaneous-move game pictured in Table 1.[13] Here the proposer chooses a column and the responder a row. The responder receives the first payoff listed, the proposer the second. So, if the proposer chooses 'Demand ⅔' and the responder '[⅓, 1]' then the proposer receives ⅔ and the responder ⅓.

Table 1 **Modified Ultimatum Game**

	Demand ⅓	Demand ½	Demand ⅔
[⅓, 1]	(⅔, ⅓)	(½, ½)	(⅓, ⅔)
[½, 1]	(⅔, ⅓)	(½, ½)	(0, 0)
[⅔, 1]	(⅔, ⅓)	(0, 0)	(0, 0)

Presented in this way, fair behavior does not survive another Nash refinement: elimination of weakly dominated strategies. One might want to eliminate those strategies that occasionally do worse and never do better than an alternative strategy. In this case, [⅓, 1] weakly dominates all other strategies since it does equally well as the others against some opponents' strategies, but does better than both against Demand ⅔.

If we are concerned with the evolution of fair strategies this may not matter – weakly dominated strategies need not be eliminated by the replicator dynamics. In order to use the single-population replicator dynamics, we must think of each player as having two strategies: a proposal strategy and a minimum amount to accept. We will represent these with the ordered pair <*a*, *b*>, where *a* is the proposal and *b* is the minimum acceptable. Each player receives the expected return from playing half the time as the proposer and half the time as the receiver against the population. In this expanded game, there are three symmetrical pure-strategy Nash equilibria (Nash equilibria where both players play the same strategy with a probability of 1). Each corresponds to a Nash equilibrium of the sequential-move ultimatum game. They are <⅔, ⅓>, <½, ½>, and <⅓, ⅔>. A population composed entirely of <⅔, ⅓> players is evolutionarily stable – any mutation will be eliminated by the dynamics. Both <½, ½> and <⅓, ⅔> are neutrally stable – some mutations will remain, but none will invade the population.

Unfortunately, the fair strategy equilibrium is not as stable as the unfair one. If the population drifts too far from being composed completely by one strategy, it can then be invaded. For illustration consider the dynamics between only three strategies: <½, ⅓>, <½, ½>, and <⅔, ⅓> (as pictured in Figure 1).[14] All populations along the (<½, ½>, <½, ⅓>) line are equilibria (since the only difference between the strategies is their response to another strategy which is not present). However, those points on the line to the left of point *a* can be invaded. Consider a population with 24 percent playing <½, ½> and 76 percent playing <½, ⅓> (past point *a*). In this population, a <⅔, ⅓> mutant does slightly better than the population and thus could invade. Moving a population with a substantial number of <½, ½> players down the line of equilibria is easier than it might seem – it does not require completely random drift to pass the critical point. Suppose a mutant <⅔, ⅓> player arises in a population composed of mostly (but not entirely) <½, ½> players. Although our mutant will be eliminated, her presence temporarily helps the <½, ⅓> players. The appearance of this mutant has moved our population down along the line of equilibria, making invasion more likely. This shows the sort of equilibria that comprise the line to the right of point *a*. Populations *can* converge toward those points, but not every population in the neighborhood of a point converges toward it. A population will not return to its prior state if invaded by a mutant, rather it will return to a nearby, similar state. The only exception is a population composed entirely of <½, ½> players; it will return exactly to that state if a <⅔, ⅓> mutant arises. However, since most

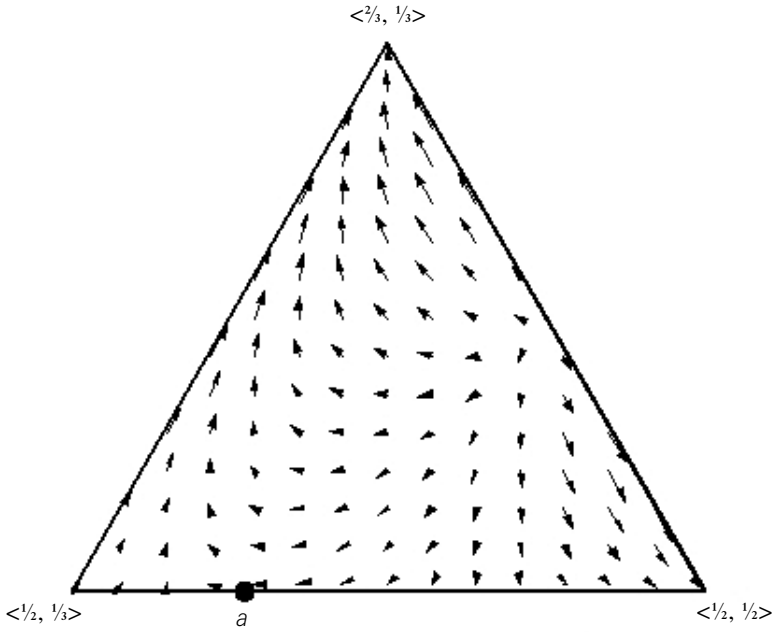

Figure 1 **Three strategies in the Ultimatum Game**

interior populations converge toward the line, not toward the vertex, we should not expect to find many populations of this type. Other than points along the line, only <⅔, ⅓> is a stable attractor. A population composed of 50 percent <½, ½> players and 50 percent <⅔, ⅓> players is an equilibrium, but a mutation in any direction will drive the population away from it.

In the larger system with all nine strategies present, there are other similar neutrally stable polymorphisms. Certain combinations of all three demand-⅓ strategies are neutrally stable – so long as a sufficient number of <⅓, ⅔> players are present, no one who demands more can invade. In a computer simulation, some populations converged toward one of these demand-⅓ polymorphisms, although the total number was less than 1 percent of the starting populations. Some 65 percent of populations converged toward a population in which the players play the weakly dominant strategy < ⅔, ⅓>. The remaining populations (34 percent) went to one of the stable (<½, ⅓> or <½, ½>) polymorphisms.[15] As suggested earlier, if we were to introduce mutations, we should expect that the long-run play would converge toward the weakly dominant equilibrium.

If we are to explain the evolution of fair behavior in the ultimatum game, we must add some additional complexity to the model. Gale, Binmore, and Samuelson do just this.[16] In their evolutionary model, there are two populations: a pop-

ulation of proposers and a population of responders. If the probability of mutation is higher in the population of responders than it is in the population of proposers, then a non-sequentially rational strategy can evolve. While this is certainly an interesting result, in order to explain the evolution of fair proposals and rejections in actual human cultures, it must be the case that people are neatly divided into two groups. Furthermore, it is unclear why we ought to expect experimentation, modeled as mutation, to be more likely in populations of responders than in populations of proposers.

Güth and Yaari present a single-population model in which fair proposals can evolve.[17] In their model, individuals are capable of recognizing their opponent's type – a questionable assumption since anonymity is maintained in most experimental settings, and yet fair behavior still remains. Huck and Oechssler relax this assumption slightly.[18] Their players are aware of the proportion of individual types in the population and then determine their proposal accordingly. For a sufficiently small population, fair behavior is the only evolutionary stable strategy. In this model, the populations must be small enough that the rejection of unfair behavior harms unfair proposers sufficiently to prevent their invasion. Again, one might worry about both of these assumptions.

Harms provides the most plausible analysis of a single-population model of the ultimatum game.[19] While his primary concern is maintaining the stability of fair behavior under mutations, he does have a model in which fair behavior can evolve. Harms considers a strategy set that contains many more divisions near ½ than divisions above ½. With this modification of the game, many more populations converge toward a near-fair division (60–70 percent). In his simulations, Harms only studies an even initial distribution over the strategies, which may tip the balance in favor of fair strategies since there were more strategies near the fair equilibrium.

3. Norms and the ultimatum game

Perhaps the explanation for fair behavior in the ultimatum game is not to be found in the properties of this game alone. Several suggestions are available for explaining the anomalous behavior. Perhaps the experimental subjects are risk averse. Players who are afraid that some unfair offers might be rejected might be willing to take a slightly lesser payoff in order to ensure that they receive something rather than nothing. Unfortunately, this model cannot account for the occurrence of actual rejections. Since the responder faces no risk, he is choosing between something and nothing, yet still sometimes takes nothing. In addition, Henrich et al. calculate the degree of risk aversion needed to account for the behavior of their experimental subjects; they determine that mere risk aversion cannot account for proposers' behavior. Furthermore, Henrich et al. examine another game in which the responder cannot decline (the dictator game), simply receiving the amount allocated to them by the proposer. While the mean amount

offered declined, it was not zero. This action cannot be explained by risk aversion since there is no risk of refusal.

Perhaps the players have a notion of fairness which modifies the payoff structure. Accepting a small positive dollar offer may have a negative expected utility for the responder since the negative utility of accepting an unfair offer might outweigh the positive utility of the money. Maybe the responders receive some benefit from punishing, or accepting a low offer may psychologically harm the responder by making her feel subordinate. Additionally, the proposer might gain some psychological benefit from acting in a fair way, so high a benefit that it exceeds the benefit gained from receiving more money. There is some evidential support for this explanation. Oosterbeek et al. find that as the stakes increase, the mean proposal decreases. Again, however, it rarely reaches the sequentially rational strategy. Although, one might be inclined to believe that something like a subjective utility function is operative in this context, it does not offer a completely satisfying explanation.

> Appeals to norms of fairness, however, hardly constitute an explanation in itself. Why do we have such norms? Where do they come from? If they are modeled as factors in a subjective utility function, how do such utility functions come to be so widespread? . . . Perhaps punishing behavior could be explained by generalization from some different context. But even if that were the case, we would still be left with the evolutionary question: Why have norms of fairness not been eliminated by the process of evolution?[20]

Here Skyrms offers one suggestion for this deeper explanation: perhaps the norm evolved in some wider context which includes interactions other than the ultimatum game.[21] Skyrms is not the only author to have suggested a general norm for the explanation of cooperation in the ultimatum game. Gale et al. suggest a very specific context in which our norms may have evolved.

> In particular we suggest that initial play reflects decision rules that have evolved in real-life bargaining situations that are superficially similar to the Ultimatum Game. These bargaining games generally feature more symmetric allocations of bargaining power than the Ultimatum Game, yielding initial play in the Ultimatum Game experiments that need not be close to [sequentially rational play].[22]

It is clear that without mentioning it specifically, Gale et al. are considering a game such as the Nash bargaining game as the situation in which norms of fairness have evolved. Certainly, they agree that occasionally we play the ultimatum game, but presumably this occurs less often than its more symmetrical counterpart.

Skyrms and Gale et al. both presume that a single norm of fairness evolved to cover both the ultimatum game and Nash bargaining game, and that this norm evolved to encourage both fair proposals and, perhaps, rejections of unfair offers. There are two questions which must be addressed in order to provide a satisfactory explanation along this line. First, why is there only a single norm of fairness,

and not a fairness-in-Nash norm and a fairness-in-ultimatum norm? Second, why would a single norm evolve to prefer the even split instead of another bargaining solution?

Regarding the first, individuals may simply fail to consider the full strategic situation with which they are confronted. Costa-Gomes, Crawford, and Broseta provide an extensive study demonstrating that a reasonable number of individuals simply do not consider game-like situations strategically.[23] In their study, many individuals only considered their own payoffs without ever investigating the payoffs of the other player and thus could not have been making informed predictions about the other player's actions. Even for those players who would consider the actions of another, in many bargaining situations the information may be strictly unavailable to the players (or at least unavailable at a reasonable expense). In the Nash bargaining game, one can pre-commit oneself to a minimum. The possibility of pre-commitment is what makes the fair equilibrium sequentially rational in the Nash bargaining game, and its absence makes fairness irrational in the ultimatum game. Situations in which individuals may have external commitments to some portion of the good provide a position in which the amount of pre-commitment is unknown to both parties. For example, a seller of a house may have already agreed to a purchase price for a new house, and thus have already effectively committed himself to obtaining a certain minimum for the house he is selling. If the buyer is unaware of this amount, or unaware if the seller has even agreed to purchase a new house, the extent of pre-commitment is unknown.

In the version of the game discussed above, in which pre-commitment is structurally part of the game, the only difference between the Nash bargaining game and the ultimatum game is what happens to the excess good when the demands (or demand and minimal acceptable amount) do not total the entire good.[24] In the Nash bargaining game the excess is wasted, but in the ultimatum game the excess is given to the respondent.

Why should we think that the players would not have information about where the excess would go? Consider two bargainers who are operating via an intermediary. If the bargainers do not know the intentions of the intermediary, they may not know what the intermediary will do with an excess should it appear. The intermediary might take the excess good for herself, effectively making the bargaining game equivalent to the Nash bargaining game, or she might give the excess to one of the bargainers, making the game the simultaneous-move ultimatum game.[25]

This topic certainly deserves more treatment than is offered here. Instead, however, we will focus on the second question. Namely, given that a single norm of fairness evolved to cover both games, why did the fair split evolve as the norm that governs both games? In assuming that the presence of the Nash bargaining game might help to drive the evolution of fair behavior in the ultimatum game, Skyrms and Gale et al. have tacitly relied on the assumption mentioned in the

Introduction – that the probability of fair behavior growing to fixation in a population lies somewhere in between the probabilities in each game individually. If their suggestions for an generalized fairness norm are to provide plausible explanations of fairness in the ultimatum game, it must be the case that the probability of fair behavior in an environment including both games is closer to 62 percent than 34 percent.

One can model this suggestion by combining the Nash bargaining and ultimatum games into a larger game of incomplete information. Since our fairness norm does not distinguish between the different games, we can treat it as playing a strategy in this larger game. With this model, we can then determine if the norm would be able to evolve and the probability of its evolution.

4. A complex environment

We will combine these games by having players adopt a single demand for the Nash bargaining game and the ultimatum game and also adopt a separate minimum acceptability threshold for the ultimatum game.[26] In order to restrict the number of strategies, we will limit our model as before. The pie is only divisible into three chunks: ⅓, ½, and ⅔. Individuals must chose a single demand for both the Nash bargaining game and the ultimatum game. In addition, individuals must adopt a minimum acceptable threshold for the ultimatum game. Nature chooses which game will be played and which player acts first, if the ultimatum game is chosen.

Suppose that nature chooses the Nash bargaining game with a probability of 1 – p and the ultimatum game with a probability of p. If the ultimatum game is chosen, nature chooses a proposer at random. We can then find the Nash equilibria of the game where payoffs are the agent's expected payoffs given p.[27] We will represent the pure-strategy Nash equilibria as unordered pairs (a, b), where a and b are the two strategies played in the game, a by one player and b by another. Unsurprisingly, the pure-strategy Nash equilibria depend on p; they are presented in Figure 2. The bars represent the values for p where the listed strategies are Nash equilibria. For all p, (<½, ½>, <½, ½>) is preserved as an equilibrium of the game. However, so are many unfair equilibria.

Only a few of the Nash equilibria are evolutionary stable. Any population composed of <½, ⅓> and <½, ½> is neutrally stable for $p < 6/7$. (As we already know, some populations composed of these two strategies are neutrally stable for all p.) A population composed entirely of < ½, ½> is evolutionary stable for all p. In addition, these populations cannot drift away from fair equilibria in the same way that populations could in the ultimatum game. However, populations that have unfair or hyper-fair (larger than ½) proposals are also evolutionarily stable.[28]

To determine the quality of explanation offered by our new model we should find the size of the basins of attraction for the different evolutionary stable states. The basin of attraction for the fair equilibria with $p = 0$ (the Nash bargaining

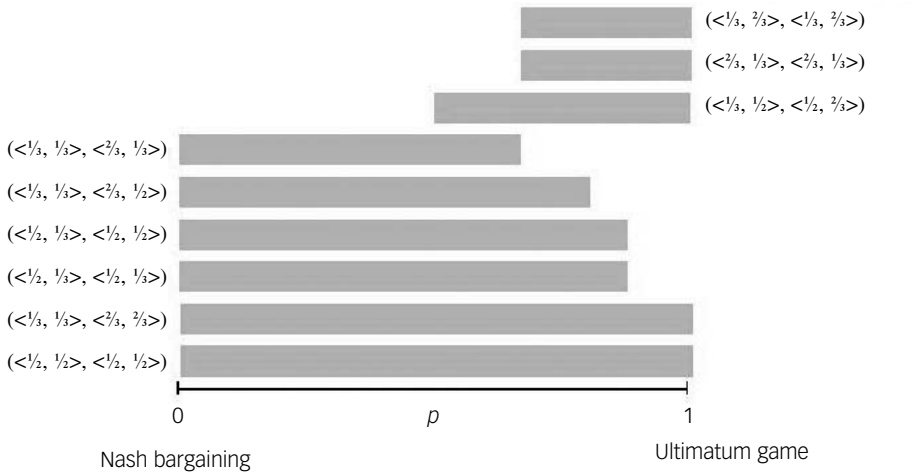

Figure 2 **Nash equilibria in terms of *p***

game) is 86 percent.[29] Setting $p = ½$, we find that the basin of attraction for fair proposers is 93 percent! The basins of attraction for fair proposals are represented in terms of p in Figure 3 (for comparison, the dotted line represents a linear combination of the ultimatum and Nash bargaining's basins of attraction). Intuitively, one would think the size of the basin of attraction for the combined game would be somewhere in between the size for each game individually, but this is not the case.

While explaining exactly why this result occurs requires a rather detailed discussion of the game and the dynamics, there are some facts that may make this seem intuitive. In the ultimatum game, a player who plays <⅔, ⅓> is at no disadvantage when meeting himself compared to a <½, ½> player. Since a ⅔ proposer in the Nash bargaining game does not do well against himself, even the slightest introduction of the Nash bargaining game results in the <½, ½> player doing better against himself.

As an illustration, consider the same three strategies present in Figure 1 (<½, ⅓>, <½, ½>, and <⅔, ⅓>) and let $p = ½$. Now the entire line between <½, ⅓> and <½, ½> is an attractor. While <⅔, ⅓> has a small basin of attraction (in this limited context), it represents only a small portion of the simplex. In fact, <⅔, ⅓> is not globally stable in the whole population.

It is not as obvious why we ought to expect the basin of attraction to increase over the Nash bargaining game. To understand why this result occurs, we will look at the evolution of one population over time. All of the starting populations that evolve toward the unfair equilibrium of the Nash bargaining game start out

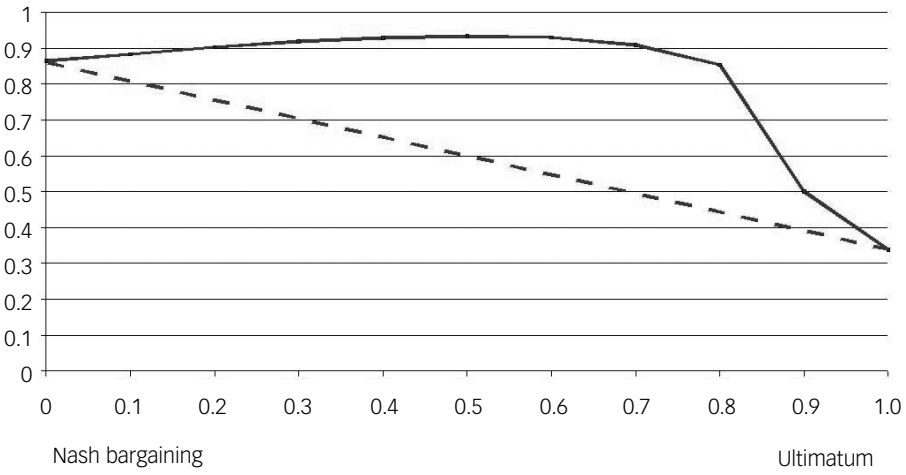

Figure 3 **Basins of attraction of fair proposers in terms of *p***

with relatively high proportions of ⅓ and ⅔ proposers. Consider the population proportions in Table 2. This table represents one state of the population and compares the relative payoffs for each strategy in two games: the pure Nash bargaining game and the game that combines the Nash bargaining and ultimatum games with equal probability. Here, there are two important differences between the Nash bargaining game and the mixed game. In the Nash bargaining game, these initial proportions help all ⅓ proposers most and then also help all ⅔ proposers. In the combined game, the very timid players (<⅓, ⅓>) are helped most of all, followed by the timid fair proposers (<½, ⅓>). A graph of the evolution of these two strategies over time is presented in Figure 4. Once ⅓ and ½ proposers compose a substantial part of the population, ½ proposers do much better than ⅓ proposers. As a result, they grow to take over the population.

In this model, we only have a limited explanation for the rejection of unfair offers. Very few of our populations resulted in an end state that was entirely composed of <½, ½>. Usually, the end population contained both <½, ⅓> and <½, ½>. This is not an implausible result since in experiments some unfair proposals are accepted.

This model nicely coincides with more specific experimental results as well. Fischer et al. have conducted an experiment on a game strategically similar to ours.[30] In their experiment, individuals were assigned to be either first or second players. In their game, the first player makes a proposal and with some fixed probability, *p*, the second is told the proposal amount. If the second is told the amount, she can choose either to accept or reject (she is playing the ultimatum game). If the second is not told, she must make a demand, and if the demands are

Table 2 **Payoffs for one population in two games**

Strategy	Proportion of the population	Payoff in Nash bargaining game	Payoff for $p = ½$
<⅓, ⅓>	0.17	0.33	0.38
<⅓, ½>	0.17	0.33	0.34
<⅓, ⅔>	0.17	0.33	0.33
<½, ⅓>	0.05	0.28	0.35
<½, ½>	0.01	0.28	0.31
<½, ⅔>	0.04	0.28	0.3
<⅔, ⅓>	0.05	0.31	0.3
<⅔, ½>	0.17	0.31	0.29
<⅔, ⅔>	0.17	0.31	0.28

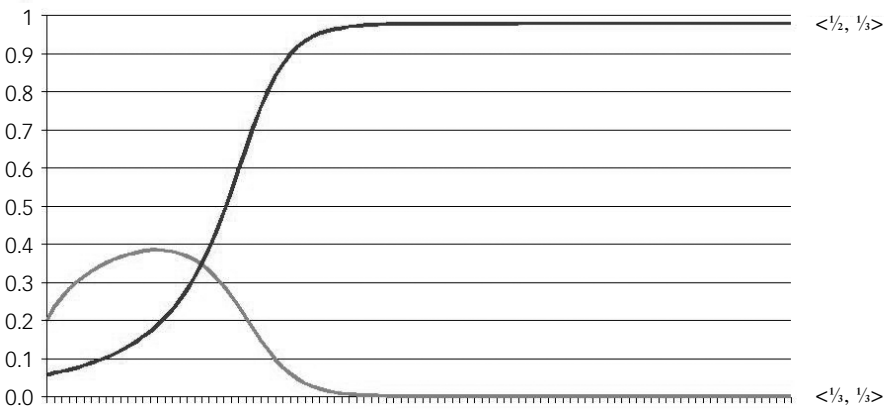

Figure 4 **Two strategies over time**

compatible both players get their demand (the Nash bargaining game). Fischer et al. studied the same subjects playing for different values of p, where the subjects were aware of the value of p for each round. They found that the modal offer was 50 percent and the mean varied from 50.32 percent to 58.61 percent (as p varied from 0.1 to 0.9). Also, they found that over time their subjects *increased* the number of fair proposals, suggesting the results from our evolutionary model nicely coincide with individuals' learning behavior in the game.

5. Conclusion

This result provides a potential explanation for cooperative behavior in the ultimatum game. In addition, we have an enhanced explanation for cooperative behavior in the Nash bargaining game, since fair populations in the combined game have larger basins of attraction than the Nash bargaining game alone. Not only has this model provided interesting results on its own, but it also suggests a fruitful avenue of research for modeling norms. It certainly seems plausible that people do not process all the strategic details of every situation with which they are confronted. Even if it were possible, in many circumstances the costs might outweigh the benefits of doing so. Given that people use heuristics for a large class of games, this model provides an evolutionary explanation for the emergence of a norm of fairness in bargaining.

This model also suggests that the explanation for particular human behaviors may reside outside the features of any individual game. Rather, the interaction of several different strategic situations may result in unexpected outcomes. This suggests that in seeking evolutionary explanations, we should not limit ourselves to studying one game in isolation. While interesting explanations can emerge from studying a single game, we have not exhausted all potential explanations for social behavior by studying it alone. Finally, we should not uncritically generalize an explanation from simple games to explanations for human behavior. In order to use a simple game as an explanation for human behavior, we must have some justification of our assumption that the increased complexities present in actual human experience do not radically alter the properties of the system.

notes

Thanks to Brian Skyrms, the anonymous referees, and the participants of UCI's dynamics seminar for their very helpful comments on earlier drafts. This research was conducted with generous financial support from the Institute for Mathematical Behavioral Science and the School of Social Science at UCI.

1. For discussion of the relationship between justice and the Nash bargaining game, see Jason Alexander and Brian Skyrms, 'Bargaining with Neighbors: Is Justice Contagious?' *Journal of Philosophy* 96 (1999): 588–98; Ken Binmore, *Game Theory and the Social Contract Volume 2: Just Playing* (Cambridge, MA: MIT Press, 1998); Ken Binmore, *Natural Justice* (Oxford: Oxford University Press, 2005).
2. For evolutionary explanations of cooperative behavior in the Nash bargaining game, see Jason McKenzie Alexander, 'Evolutionary Explanations of Distributive Justice', *Philosophy of Science* 67 (2000): 490–516; Brian Skyrms, *Evolution of the Social Contract* (Cambridge: Cambridge University Press, 1996); Brian Skyrms, 'Signals, Evolution and the Explanatory Power of Transient Information', *Philosophy of Science* 69 (2002): 407–28; H. Peyton Young, *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions* (Princeton, NJ: Princeton

University Press, 1998). For the ultimatum game, see John Gale, Kenneth G. Binmore and Larry Samuelson, 'Learning to be Imperfect: The Ultimatum Game', *Games and Economic Behavior* 8 (1995): 56–90; Werner Güth and Menachem Yaari, 'An Evolutionary Approach to Explain Reciprocal Behavior in a Simple Strategic Game', in *Explaining Process and Change: Approaches to Evolutionary Economics*, edited by U. Witt (Ann Arbor: University of Michigan, 1995), pp. 23–34; William Harms, 'Evolution and Ultimatum Bargaining', *Theory and Decision* 42 (1997): 146–75; Steffen Huck and Jörg Oechssler, 'The Indirect Evolutionary Approach to Explaining Fair Allocations', *Games and Economic Behavior* 28 (1999): 13–24.

3. While there have been several authors who have provided models that predict fair behavior in the ultimatum game, many of these models rely on unrealistic assumptions. They will be discussed in more detail in Section 2.
4. There is also a Nash equilibrium where both players demand the entire good. Recall that in order to be a Nash equilibrium it must be the case that neither player would do better by *unilaterally* changing his strategy. Since an individual player reducing his demand without cooperation of the other would not increase his payoff, this combination of strategies constitutes a Nash equilibrium.
5. Rudy V. Nydegger and Houston G. Owen, 'Two-Person Bargaining, an Experimental Test of the Nash Axioms', *International Journal of Game Theory* 3 (1974): 239–50; Al Roth and Michael Malouf, 'Game Theoretic Models and the Role of Information in Bargaining', *Psychological Review* 86 (1979): 574–94; John van Huyck, Raymond Battalio, Somesh Mathur, P. van Huyck and Andreas Ortmann, 'On the Origin of Convention: Evidence from Symmetric Bargaining Games', *International Journal of Game Theory* 34 (1995): 187–212; Menachem Yaari and Maya Bar-Hillel, 'On Dividing Justly', *Social Choice and Welfare* 1 (1981): 1–24.
6. Informally, a state is evolutionarily stable if it cannot be invaded by a mutant strategy. For a formal definition, see John Maynard Smith, *Evolution and the Theory of Games* (Cambridge: Cambridge University Press, 1982).
7. Skyrms, *Evolution of the Social Contract*; Skyrms, 'Signals, Evolution and the Explanatory Power of Transient Information'.
8. See Alexander, 'Evolutionary Explanations of Distributive Justice'; Skyrms, 'Signals, Evolution and the Explanatory Power of Transient Information'; Young, *Individual Strategy and Social Structure*. However, adding realistic assumptions does not guarantee an increase. One study found that under one realistic modification, the basin of attraction of fair behavior shrinks. See Justin D'Arms, Robert Batterman and Krzyzstof Gorny, 'Game Theoretic Explanations and the Evolution of Justice', *Philosophy of Science* 65 (1998): 76–102.
9. There is also another Nash equilibrium where the first player proposes to keep the entire good and the second rejects all proposals which give her less than the entire good. The second would still receive nothing by accepting and the first cannot propose any split that the second would accept (and would result in her receiving more than zero).
10. See Reinhard Selten, 'Reexamination of the Perfectness Concept of Equilibrium in Extensive Games', *International Journal of Game Theory* 4 (1975): 25–55.
11. Hessel Oosterbeek, Randolph Sloof and Gijs van de Kuilen, 'Cultural Differences in

Ultimatum Game Experiments: Evidence from a Meta-Analysis', *Experimental Economics* 7 (2004): 171–88.

12. See Joseph Henrich, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr and Herbert Gintis, *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies* (Oxford: Oxford University Press, 2004); Joseph Henrich, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, Herbert Gintis, Richard McElreath, Michael Alvard, Abigail Barr, Jean Ensminger, Kim Hill, Francisco Gil-White, Michael Gurven, Frank Marlowe, John Q. Patton, Natalie Smith and David Tracer, '"Economic Man" in Cross-cultural Perspective: Behavioral Experiments in 15 Small-Scale Societies', *Behavioral and Brain Sciences* 28 (2005): 795–855.
13. This assumption has already limited our ability to explain all the data on the ultimatum game. Henrich et al. observe that some hyper-fair offers (that is, larger than ½) are rejected in some societies. Since this is a relatively rare behavior that Henrich et al. suggest can be explained by a peculiar feature of a few cultures, I suspect its explanation resides outside of an explanation for more robust irrational behavior. See Henrich et al., *Foundations of Human Sociality*; Henrich et al., '"Economic Man" in Cross-cultural Perspective'. As an interesting aside, Oosterbeek et al. find that when the game is described this way in experiments the rejection rate is significantly higher. See Oosterbeek et al., 'Cultural Differences in Ultimatum Game Experiments'.
14. This represents one face of the nine-dimensional simplex.
15. The simulation results are for the standard discrete time replicator dynamics whereby a type's frequency in the next generation is determined by its previous frequency and its payoff against the population. These simulation results are without mutation. A similar game which includes strategies that reject hyper-fair proposals is analyzed in Skyrms, *Evolution of the Social Contract* and Harms, 'Evolution and Ultimatum Bargaining'. Their results also show that <½, ½> can be evolutionarily stable, but its basin of attraction is relatively small.
16. Gale et al., 'Learning to be Imperfect'.
17. Güth and Yaari, 'An Evolutionary Approach to Explain Reciprocal Behavior in a Simple Strategic Game'.
18. Huck and Oechssler, 'The Indirect Evolutionary Approach to Explaining Fair Allocations'.
19. Harms, 'Evolution and Ultimatum Bargaining'.
20. Skyrms, *Evolution of the Social Contract*, p. 28.
21. The strategy of employing several different games in order to explain the emergence of norms governing behavior has been recently pursued by Bednar and Page. They study several common 2×2 games allowing players to adopt similar strategies across games. See Jenna Bednar and Scott Page, 'Can Game(s) Theory Explain Culture? The Emergence of Cultural Behavior within Multiple Games', Santa Fe Institute Working Paper 2004-12-039 (Santa Fe, NM: Santa Fe Institute, 2004).
22. Gale et al., 'Learning to be Imperfect', p. 59.
23. Miguel Costa-Gomes, Vincent P. Crawford and Bruno Broseta, 'Cognition and Behavior in Normal-Form Games: An Experimental Study', *Econometrica* 69 (2001): 1193–235.
24. This observation was first made by Roberto A. Weber, Colin F. Camerer and Marc

Knez, 'Timing and Virtual Observability in Ultimatum Bargaining and "Weak Link" Coordination Games', *Experimental Economics* 7 (2004): 25–48.

25. I am indebted to Brian Skyrms for this suggestion.
26. We might have combined the game by forcing a player to adopt one value which functions as both a demand and as a rejection threshold. Our method seems more intuitive since a player may be making a demand unaware if the other is simultaneously making a demand or not, but not unaware that she is in a position of accepting and rejecting. This also allows us to analyze a situation in which the norm can maximally distinguish between the two games without adopting completely different strategies for each. In fact, this method is confirmed by the analysis of experimental subjects' strategies in a similar game studied by Sven Fischer, Werner Güth, Wieland Müller and Andreas Stiehler, 'From Ultimatum to Nash Bargaining: Theory and Experimental Evidence', unpublished manuscript, 2003.
27. Strictly speaking, the Nash equilibrium of the expected return game is not a Nash equilibrium of the game as described. The Nash equilibrium of the expected return game is a Bayesian Nash equilibrium of the game as described. For simplicity's sake, I will use the two terms interchangeably here. For detailed discussion, see John Harsanyi, 'Games with Incomplete Information Played by Bayesian Players', *Management Science* 14 (1967): 159–82.
28. For $p \leq \frac{2}{3}$, ($<\frac{1}{3}, \frac{1}{3}>$, $<\frac{2}{3}, \frac{1}{3}>$) is the ESS, while for $p > \frac{2}{3}$, the ESSs are ($<\frac{2}{3}, \frac{1}{3}>$, $<\frac{2}{3}, \frac{1}{3}>$) and ($<\frac{1}{3}, \frac{2}{3}>$, $<\frac{1}{3}, \frac{2}{3}>$).
29. The reader may have noticed that this contradicts the 62 percent estimate for the basin of attraction in the Nash bargaining game obtained by Skyrms. This is a result of the sampling procedure. When $p = 0$, we are still sampling over a nine-dimensional space of initial populations. A population drawn from a uniform distribution over the points in this space will not be a uniform distribution over populations in a reduced three-dimensional space. One might worry that something like this fact is driving the increased basin of attraction for fair behavior in my model. While this may be a concern, there is little to be done – the ultimatum game cannot be expressed as a symmetrical 3×3 game. No matter the actual size, the relative fact remains that the combined game has a larger basin of attraction than the Nash bargaining game.
30. See Fischer et al., 'From Ultimatum to Nash Bargaining: Theory and Experimental Evidence'.