Evolution, Fairness, and the Ultimatum Game

Kevin J.S. Zollman

'I think justice ought to be fair'
-- George W. Bush (12/15/04)

In the evolutionary study of justice, two games have predominated the literature the Nash Bargaining Game and the Ultimatum Game. Both games provide an interesting context in which to study fair behavior, a central feature of our conceptions of justice.[1] In addition to considerations of fairness, the Ultimatum Game allows us to analyze costly punishment for other's unfair behavior. While the evolution of fair play in the Nash Bargaining game have been modeled somewhat extensively,[2] the Ultimatum Game has more problems. All Nash refinement criteria have unique predictions in the Ultimatum Game, but those predictions are rarely observed. While this does not render Nash refinements intellectually bankrupt, it provides a strong challenge to Nash refinements as both predictive tools and explanatory devices. Even evolutionary accounts have had limited success in explaining fair behavior. Here I will present a model that attempts to account for the observed behavior. This is achieved by limiting information that is normally believed to be available to the experimental subjects. It is possible that while the information is consciously available to the subjects they may have an overwhelming urge to play a certain way based on a cultural (or biological) norm that does not make the relevant distinction. Using standard Nash Refinement criteria, this model fairs only slightly better than the standard model. However, using an evolutionary approach this model holds great promise as an explanation of fair behavior.

1 See (Alexander and Skyrms 1999) and (Binmore 1998) for discussion of the Nash Bargaining Game's relationship to justice.

2 Jason Alexander (2000) and Brian Skyrms (1996; 2004) discuss one approach for explaining cooperative behavior in the Nash Bargaining Game.

The Ultimatum Game

The Ultimatum Game has two players. The first proposes a split of a good. The second, knowing the offer, either accepts or rejects the offer. If the second accepts, each player receives the proposed amount. If the second rejects, they both receive nothing. This game has several Nash Equilibria where the proposal is as beneficial for the first player as the second will accept. However, only one of these Nash Equilibria is Subgame Perfect. A strategy is Subgame Imperfect if at any point during the game a player acts in a way that results in her receiving a lower payoff than another available option. If the proposer offers a split which gives the second any positive amount, the second does strictly worse by refusing the offer. 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 the Subgame Perfect equilibria. In fact, players usually offer more than the smallest possible offer and low offers are occasionally rejected.[3] These experimental results draw into question the power of Subgame Perfection as both a predictive and explanatory tool.

In an extensive cross-cultural study, Henrich, et al. (forthcoming) observed a wide variety of strategies employed in the Ultimatum Game. While some small scale cultures did appear to play the Subgame Perfect equilibria, many did not. Henrich, et al. also observed that one's play in the Ultimatum Game is correlated with one's culture and not

3 Oosterbeek, Sloof, and van de Kuilen (2004) provide a nice overview of the experimental literature. They perform an analysis on several datasets and find that the average offer to the second player was 40% of the good and 16% of offers were rejected.

with one's own standing in that culture.

These results indicate something very important. Behavior in the Ultimatum Game is culturally contingent and usually not Subgame perfect. Furthermore, individual behavior in the Ultimatum Game is governed by one's culture more than any feature about oneself. The important challenge raised by Henrich's et al. experiment, is to construct a model that allows Subgame Imperfect play to evolve, but its evolution depends on particular features of the environment in which the norm evolves.

Several attempts have been suggested to save Nash Refinements (like Subgame Perfection) in light of the experimental results. First, one might imagine that the players have a notion of fairness which modifies the payoff structure. So, a positive dollar offer may have a negative expected utility for the players. Perhaps the players believe they receive some benefit from maintaining a social norm, or perhaps accepting a low offer may psychologically harm the second player by making them feel subordinate. Although this certainly seems a likely candidate for explaining the behavior, it is not completely satisfying. Within a rational choice context, simply appealing to a subjective utility function is tantamount to abandoning rational choice explanations. One's subjective utility function need not be constrained by anything like “rationality.” As a result, supposing that people are rationally maximizing an irrationally constructed utility function would hardly constitute a victory of the rational choice paradigm.[4]

A more palatable version of this solution is to suggest that players are risk-averse. One this suggestion, players are afraid that some unfair offers might be rejected and so are willing to take a slightly lesser payoff in order to ensure that they receive something

4 Bethwaite and Thompson (1996) attempt this sort of explanation.

rather than nothing. Unfortunately, this model cannot account for the occurrence of actual rejections since it would still be better to accept something rather than 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 individuals behavior. (Henrich, et al. forthcoming, 17-18)

The second option is that players do not process all the strategic details of the game. They recognize certain features of a game as conforming to a general structure and then act on the basis of a norm governing games of that structure. Of course, there are many norms that might govern behavior over several games, and mere appeal to a social norm as governing behavior fails to provide a satisfactory explanation.

> Appeals to norms of fairness, however, hardly constitutes 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? (Skyrms 1996, 28)

This paper presents a model that offers some hope at providing a deeper explanation of the second sort. The fundamental tenet of this model is that a norm is a strategy for several different (but similar) games and this norm does not have a game contingent strategy. This can be modeled game theoretically by combining several simple games into a larger game of incomplete information. Since the norm does not distinguish between the different games, we can treat it as playing a strategy in this larger game of incomplete information.[5] With this model, we can then determine if the norm would be

5 We need not consider the information as being strictly *unavailable*, but rather unused in the strategic calculation. This situation may occur for any number of reasons: the information may be unavailable, the agent may have not considered the information relevant, obtaining the information may be so costly

able to evolve.

This model is not presented as *the* model for all Ultimatum Game behavior. The importance of Henrich's et al. experiment has been to demonstrate that different behavior has evolved in different contexts. Rather, this model is offered as a potential explanation for some pro-social behavior.

Evolution of Norms

Several evolutionary accounts have been offered in the literature. Güth and Yaari (1992) present a model where fair proposals often evolve. 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 (1999) relax this assumption slightly. Their players are aware of the proportion of individual type in the population and then determine their proposal accordingly. For 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.

Following Skyrms (1996), we might think of strategies in the Ultimatum Game as a prior commitment which need not be Subgame Perfect. 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 (, ½, and) and three ranges of acceptability ([, 1], [½, 1], [, 1]), we transform the two stage Ultimatum

that the information would not be obtained, or perhaps communicable social norms cannot be sufficiently detailed.

game into a simultaneous move game (pictured in Table 1).[6]

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

Table 1: Modified Ultimatum Game

This modification removes the problem of Subgame Perfection from the standard Ultimatum Game; strategies that involve rejections of small offers are Subgame Perfect in the Modified Ultimatum Game. However, we have eliminated this problem by fiat, transforming the sequential move game into simultaneous game. In fact, refusal strategies fails another equilibrium refinement, Trembling Hand Perfection.[7] If a player believes that she is playing against another who wishes to play a certain strategy, but occasionally makes mistakes (i.e., has a trembling hand) one might play differently. In this case the best response to a trembling proposer is [, 1], since [, 1] weakly dominates all other strategies. Thus the only Nash Equilibrium which is Trembling Hand Perfect is the proposer offering and the responder accepting any offer.

In the evolutionary context this suggestion helps somewhat. We must think 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

6 This assumption has already limited our ability to explain all the data on the Ultimatum Game. Henrich, et al. (forthcoming) observe that some hyper-fair (i.e. larger than ½) offers are rejected in some societies. Since this is a relatively rare behavior that Henrich, et al, suggest can be explained by peculiar feature of a few cultures, we suspect its explanation resides outside of an explanation for more robust "irrational" behavior.

7 See (Selten 1975).

as the proposer and half the time as the receiver against the population. All three Nash Equilibria of this game are evolutionary stable. Although, two of them have interesting properties. < , > is asymptotically stable, any mutation will be eliminated by the dynamics. However, <½, ½> and < , > are neutrally stable. Some mutations will remain but none will invade the population. However, if the population drifts too far from being composed completely by one strategy, it can then be invaded. As an illustration consider a population composed with 24% playing <½, ½> and 76% playing <½, >. In this population a < , > or < , ½> mutant does slightly better than the population and will thus invade. So, while it would take substantial drift, populations *can* drift away from the Trembling Hand Imperfect equilibria in a way that cannot occur with the < , > equilibrium. In addition, the basin of attraction of <½ , ½> is relatively small. In computer simulation only 34% of the initial starting populations evolved to fair proposals.[8]

So far this modification has been of limited help. However, as suggested in the introduction, we might also modify the game by combining it with another, similar game. There are of course many different games that could be combined in an attempt to model this situation. Here we will use the Nash Bargaining game as a counterpart.[9] The Nash Bargaining Game has several strategic and heuristic similarities to the Ultimatum Game; it is not improbable that an actor might confuse the two. In addition, Nash Bargaining situations are often used as models of many commercial transactions which makes it a

8 The results in this paper are for the standard discrete time replicator dynamics where an type's frequency in the next generation is determined by its previous frequency and its payoff against the population. Skyrms (1996) and Harms (reported in Skyrms 1996) analyze a similar game which includes strategies that reject hyperfair proposals. Their results also show that <½, ½> can be evolutionarily stable, but its basin of attraction is relatively small.

9 This was initially suggested by Brian Skyrms (1996, 28).

likely standard on which individuals are basing their judgments.

For those who are unfamiliar, in the Nash Bargaining Game a player proposes a split of some divisible good (like in the Ultimatum Game). Unlike the Ultimatum Game, players *simultaneously* propose an amount of the good which they would like for themselves. If the two proposals are compatible (i.e. they do not sum to more than the total good) then each receives her demand. Otherwise both receive nothing. Like the Ultimatum Game, there are many Nash Equilibria; any two offers which sum to one are Nash Equilibria. Unlike the Ultimatum Game, all of the equilibria are Trembling Hand Perfect. Using the three possible proposals used in the Ultimatum Game, we find there are two evolutionary stable states of the population – one with fair proposals. The basin of attraction of fair proposals is relatively large, approximately 86% of the initial starting populations.

We might combine these games in two ways. First, we might make the norms maximally blind to differences in the games and have a player adopt the same demand and minimum threshold for acceptability. This seems unlikely, since players would certainly be aware if they are in the position of accepting or rejecting. Second, we might imagine that the players must 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. This seems more intuitive since a player may be making a demand unaware if the other is simultaneously making a demand or not. This also allows us to analyze a situation where the norm can distinguish between the two games as much as possible without adopting completely different strategies for each.

In order to restrict the number of strategies, we will limit our model as before.

The pie is only divisible into three chunks, , ½, and . Individuals must make a demand unaware if they are playing the Nash Bargaining Game or the Ultimatum Game. In addition, individuals must choose a minimum acceptable threshold for the Ultimatum Game. Nature choses 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 probability *(n-1)* and the Ultimatum Games with probability *n*. If the Ultimatum Game is chosen natures chooses a proposer at random. We can then find the Nash Equilibria of the game where payoffs are the an agent's expected payoffs given *n*.[10] Unsurprisingly, the Nash Equilibria depend on *n*. Figure 1 presents the Nash Equilibria in terms of *n*. Happily <½, ½> - <½, ½> (the fair equilibria with rejection of unfair offers) is preserved as an equilibria of the game for all *n*. However, so are many unfair equilibria.

Unfortunately for all this effort, <½, ½> - <½, ½> is still not Trembling Hand perfect. <½, > weakly dominates <½, ½>, because it does equally well against many of the strategies and better against any strategy that proposes . So, one would prefer <½, > to <½, ½> against any purely mixed strategy. However, whenever <½, > is a Nash Equilibrium it is Trembling Hand Perfect. <½, > is a best response against any mixed strategy that plays <½, > with at least probability 1/9 and plays each other strategy with equal probability.[11]

10 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 (Harsanyi 1967). For simplicity sake I will use the two terms interchangeably here.

11 In fact, <½, > may be a best response to other mixtures as well. Unfortunately, the complexity of determining the constraints is beyond the power of this writer and his computer.

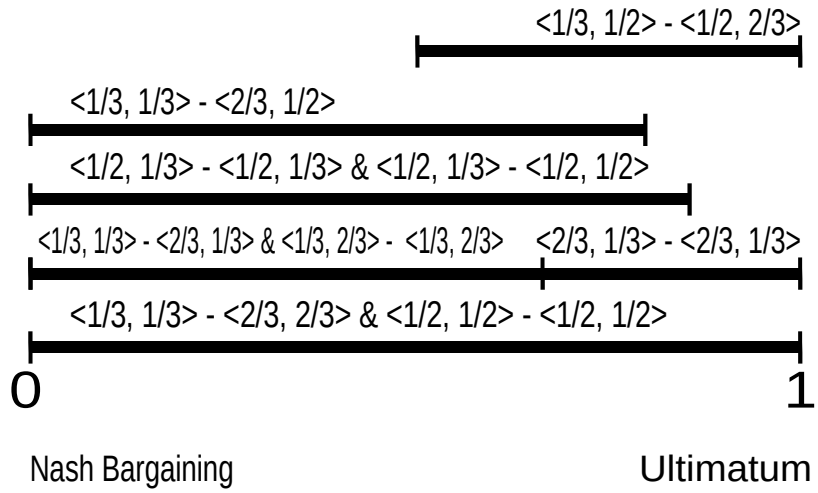

Figure 1: Nash Equilibria for combined game in terms of n

The evolutionary results are much more interesting. Only a few of the Nash Equilibria are evolutionary stable. Any population composed of <½, > and <½, ½> is stable for $n < 6/7$. A population composed entirely of <½, ½> is evolutionary stable for all n. 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 hyperfair (larger than ½) proposals are also evolutionarily stable.[12] To determine the quality of explanation offered by our new model we should determine the relative basins of attraction for the different evolutionary stable states. Recall that in the Nash Equilibrium the basin of attraction for the fair equilibria was estimated to be around 86%. Setting n=½, we find that the basin of attraction for fair proposers is 93%! The basins of attraction for fair proposals are represented in terms of n in Figure 2. This is of course a surprising result. Intuitively the 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.

12 < , > - < , > is ESS for $n <$ and < , > - < , > for $n >$.

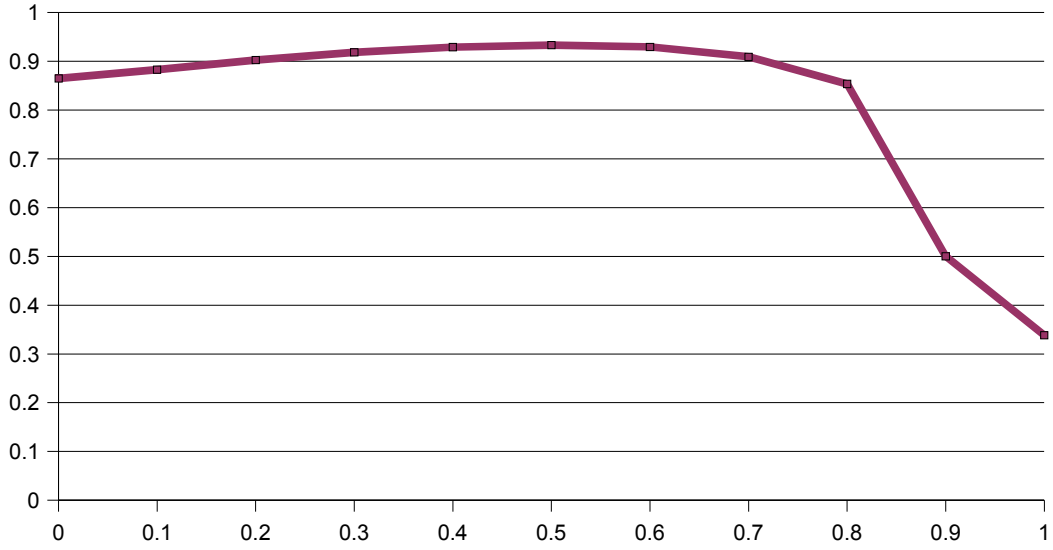

Figure 2: Basins of attraction of fair proposals in terms of n

To understand why this strange result occurs, we need to look at the evolution of one population over time. Starting populations that evolve to the unfair equilibrium of the Nash Bargaining Game start out with relatively high proportions of proposers. Consider the population proportions in Table 2. 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 proposers most and then also help all proposers. In the combined game the very timid players (< , >) are helped most of all, followed by the timid fair proposers (<½, >). A graph of the evolution of these two strategies over time is presented in Figure 3. Once and ½ proposers compose a substantial part of the population, ½ proposers do much better than proposers. As a result, they grow to take over the population.

Strategy	***Proportion***	***Nash Payoff***	***N=½***
< , >	0.17	0.33	0.38
< , ½>	0.17	0.33	0.34
< , >	0.17	0.33	0.33
<½, >	0.05	0.28	0.35
<½, ½>	0.01	0.28	0.31
<½, >	0.04	0.28	0.3
< , >	0.05	0.31	0.3
< , ½>	0.17	0.31	0.29
< , >	0.17	0.31	0.28

Table 2: Payoffs in different games

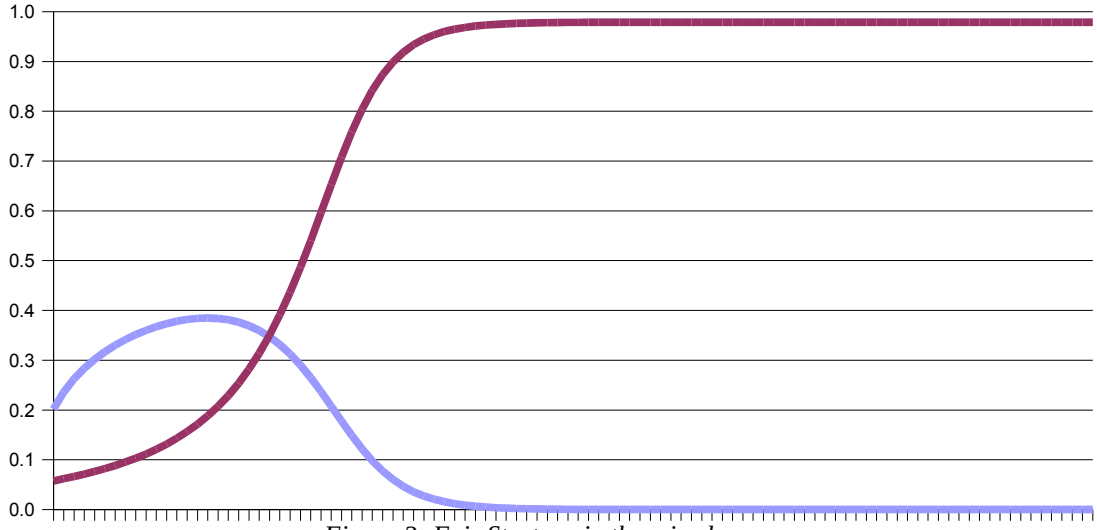

Figure 3: Fair Strategy in the mixed game

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, > and <½, ½> players in approximately equal proportions. This is not an terrible result, however, since in experiments some unfair proposals are accepted.

Conclusion

In addition to providing a potential explanation for cooperative behavior in the Ultimatum Game, this model provides a new 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 we have a better explanation for cooperation in the Nash Bargaining Game as well. Not only has this model provided interesting results on its own, but it also suggests a fruitful avenue of research for modeling norms of fairness. 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 cost 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 how a norm of fairness in the both the Ultimatum and Nash Bargaining Games might grow to fixation in a population.

References

Alexander, Jason McKenzie (2000) "Evolutionary Explanations of Distributive Justice." *Philosophy of Science* 67: 490-516.

Alexander, Jason and Brian Skyrms (1999) "Bargaining with Neighbors: Is Justice Contagious" *Journal of Philosophy* 96(11): 588-598.

Bethwaite, Judy and Paul Tompkinson (1996) "The Ultimatum Game and Non-Selfish Utility Functions" *Journal of Economic Psychology* 17: 259-271.

Binmore, Ken (1998) *Natural Justice* Manuscript

Güth, W. and M. Yaari (1992) "An Evolutionar Approach to Explain Reciprocal Bhavior in a Simple Strategic Game" in (U. Witt ed) *Explaining Process and Change – Aproaches to Evolutionary Economics* Ann Arbor, 23-34.

Harsanyi, J. (1967) "Games with incomplete information played by Bayesian players." *Management Science* 14:159-182.

Henrich, Joseph, 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" *Manuscript*

Huck, Steffen and Jörg Oechssler (1999) "The Indirect Evolutionary Approach to Explaining Fair Allocations" *Games and Economic Behavior* 28: 13-24.

Oosterbeek, Hessel, Randolph Sloof, and Gijs van de Kuilen (2004) "Cultural Differences in Ultimatum Game Experiments: Evidence from a Meta-Analysis." *Experimental Economics* 7: 171-188.

Selten, R. (1975) "Reexamination of the Perfectness Concept of Equilibrium in Extensive Games." *International Journal of Game Theory* 4:25-55.

Skyrms, Brian (1996) *Evolution of the Social Contract*. Cambridge: Cambridge Press.

Skyrms, Brian (2004) *The Stag Hunt and the Evolution of Social Structure*. Cambridge: Cambridge Press.