Final Version (February 19, 2009)

Evolutionary Considerations in the Framing of Social Norms

Brian Skyrms
U. C. Irvine and Stanford University

Kevin Zollman
Carnegie Mellon University

In the implementation of social norms, context and framing effects can make enormous differences. This is commonplace in the rich social psychology literature that is brought to bear in Cristina Bicchieri's *The Grammar of Society*. She sees the application of norms being controlled by social scenarios or *scripts* that people act out. The same problem may elicit different behavior depending on which script has been initiated and which norm it has activated.

The importance of such framing of a decision problem has been so well documented in experimental psychology and economics that experimentalists no longer spend much time trying to elicit it. Rather they take great care in constructing experiments with "neutral" protocols, designed to avoid framing the problem is one way or another.

But we really need a theory that explains both the genesis of norms and the possibility of framing effects. What might the framework for such a theory be? As David Hume saw long ago, social norms arise by a slow process of cultural evolution.

> "Nor is the rule concerning the stability of possession the less deriv'd from human conventions, that it arises gradually, and acquires force by a slow progression, and by our repeated experience of the inconveniences of transgressing it…."

Evolutionary game theory is the correct locus of our concerns.

Most evolutionary analyses concentrate on specific games. It is clear, however, that we do not evolve separate norms for individual games. Social norms evolve in the context of broad classes of social interactions – in a variety of games. Furthermore, they evolve in a way conducive to ambiguity in application. If there was a partition of social interactions, with a separate norm evolving for each class in the partition, application would be straightforward. But the classes of interactions driving the application of norms need not have any nice structure, and the norms that evolve may well not either.

A specific kind of interaction of the kind that game theorists put under the microscope – for instance, ultimatum bargaining – may be a member of various classes of social interactions. Each of these classes may carry its own norm, quite appropriate to the class, and the norms may conflict. We suggest that framing of a decision problem should be interpreted as a signal about the relevant class of social interactions, and that evolutionary analysis should be redirected to systems of classes of social interactions. In this way, many "anomalous" findings of experimental game theory may lose the air of mystery. We illustrate the approach here with two small examples, which illustrate two different types of framing effects.

Individual behavior across strategic contexts can often be very similar. For instance, it is remarkable how stable bargaining behavior is across contexts that are, from an economic perspective, strategically very different. Fair bargaining outcomes, where individuals both receive equal share of a good to be split, have been observed in many strategically different games – from the Nash bargaining game, where individuals each choose an amount to demand for themselves, to the ultimatum game, where one individual suggests a split of a good, to the dictator game, where one individual unilaterally decides how to split the good.

The later two have been viewed as especially problematic. In the ultimatum game one individual suggests a split of the good and the other is given the opportunity to accept or reject the split. If the proposer presumes that his counterpart is rational and prefers something to nothing, then the proposer expects that the responder will accept any positive offer – since she is choosing between something and nothing. With this presumption, the proposer maximizes his own return by suggesting a split which gives the responder the smallest positive amount possible. Strategy sets which survive this reasoning are known as sequentially rational equilibria, but experiments have demonstrated that often individuals do not even approximate sequentially rational play.

Even more perplexing are results from dictator games, where the option to refuse has been removed. In these games, a single individual determines a split and the second must take the offer. Here, if the first prefers more money to less, he does best by keeping all the money for himself. However, experiments again show that individuals do not behave in this way.

Turning from a traditional equilibrium analysis to an evolutionary one does not entirely remove the mystery. While fair behavior in the ultimatum game can evolve in the replicator dynamics – one model of cultural evolution – it is unstable to many types of mutation. Even considering a limited form of mutation, standard evolutionary models make evolution of fair behavior relatively unlikely.[1] In the dictator game it cannot evolve at all.

If one is to explain this phenomenon, it seems natural to suppose that individuals are relying on a single norm of bargaining when deciding what to do in many different bargaining like circumstances, and that this norm does not differentiate between different strategic contexts. This suggestion has been echoed many times, perhaps most clearly by Gale, Binmore, and Samuelson (1995),

> 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] (59).

[See also the discussion in Skyrms (1996) Ch. 2]

The evolutionary question then is "is it possible for a general norm to evolve, such that the norm results in fair behavior both in the ultimatum and Nash

1 For a detailed study of evolutionary dynamics in the ultimatum game see Binmore, Gale and Samuelson (1995) and (Harms 1997).

bargaining games?" This question was investigated by Zollman (2008). He considers a single norm which dictates both the proposal in the symmetric Nash bargaining game – where both individuals simultaneously make demands – and in the asymmetric ultimatum game. He finds that when individuals cannot distinguish between these two circumstances, fair behavior is very likely to evolve. More surprisingly, he finds that there is a synergistic effect, and that fair behavior is more likely to evolve in the generic context than it is in *either* game taken alone.

Zollman assumes that the fairness norm in question simply cannot distinguish between the two games, thus forcing it to evolve consistent behavior in both contexts. He defends this constraint by saying,

> 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... 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). (Zollman 2008, 89)

The final remark suggests that it may be too expensive for individuals to make distinctions between different forms of the game, and thus they must adopt a strategy which is not game contingent. This possibility is analyzed in detail by Mengel (2008). Mengel considered a situation where individuals confront many games, but incur some cost for distinguishing between them. She finds that depending on cost, many strategically very different situations can arise. New equilibria can be created and other equilibria destabilized. This suggests that the phenomenon found by Zollman may be a general feature of the evolution of norms.

These papers model a situation where two strategically different interactions are coalesced into a single socially relevant context for an individual. Some players may frame ultimatum bargaining as just a bargaining problem, some as a very special kind of bargaining, with different behavior as a result.

This is not the only time that framing effects matter, it can also be the case that individuals behave differently in one strategic context by cuing on some, strategically irrelevant, piece of information. For instance, it has been observed that in bargaining games apparently irrelevant bits of information can radically alter the agreed upon outcome. Although many such experiments have been performed, perhaps the most incredible results come from Mehta, Starmer, and Sugden (1992). In this experiment two individuals were randomly given four cards each from a reduced deck of cards containing only aces and deuces. Mehta, Starmer, and Sugden found that the focal point created by the cards resulted in asymmetric demands in the Nash bargaining game. They cued their subjects in such a way as to suggest that aces were valuable, and so individuals with more aces tended to demand more and symmetrically individuals with fewer demanded less. This shows people coordinating on different equilibria by cuing on strategically irrelevant information.

Let us move to a famous example discussed in the study of biological evolution in animal contests. Maynard Smith (1982) suggested that animal conflicts might be modeled using a game known as Hawk-Dove. In this game each individual would like to secure a resource (like food or a mate) and can do so by threatening the other. But if both threaten, they will fight which is the worst outcome of game.

	Hawk	*Dove*
Hawk	(0, 0)	(4, 1)
Dove	(1, 4)	(2, 2)

Table 1: Hawk-Dove

Table 1 provides the payoff matrix for this game. In this game there are two pure strategy Nash equilibria where one person plays *Dove* and the other *Hawk*. Such equilibria are attainable when there is an asymmetry available to the players (like one being designated the "row player"). In natural interactions where strategies correspond to types in the population, these equilibria are no longer attainable.

This can be illustrated with a relatively simple model of evolution known as the replicator dynamics. In the replicator dynamics individuals play with a random member of a population and reproduce according to how well they did relative to the rest of the population; those that did better than average take over larger shares of the population while those that did worse shrink. When the situation is entirely symmetric in the Hawk-Dove game evolution pushes the population toward a mixed state where some proportion are playing Hawk and some are playing Dove (see illustration 1)

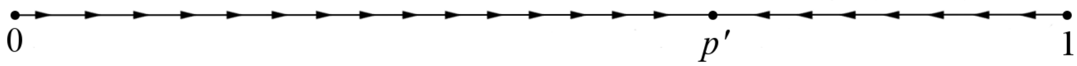

Illustration 1: One population replicator dynamics for Hawk Dove

This polymorphic equilibrium is inefficient, despite being the unique end point for this model of evolution. Hawk-types often meet other Hawk-types and have the worst payoff in the game, 0. Dove-types also meet other dove-types which results in a worse social outcome than would be secured if one had played Hawk (a sum of 4 instead of 5). Maynard-Smith noted that in such games it might be of interest to the players to find something outside of the game to use as a method for breaking symmetry. That is, if the players could use some feature observable to both players to which they can correlate their strategy, evolution might select for strategies which use this cue.

This model has been used to explain territoriality in many species. Being a "territory owner" or an "intruder" is a mechanism by which individuals might correlate their strategies, and thus solve this coordination problem. So the strategy of playing *Hawk if owner and dove if intruder* is evolutionarily stable and also more efficient than the mixed population mentioned before. We can thing of individuals sometimes being sometimes in the role of owner and sometimes in the role of intruder, with nature determining the roles with coin flips and matching owners with intruders.

Individuals now can evolve role-based strategies of the form <do this if owner, do that if intruder.>. There are four such strategies, <H,H>, <D,D>, <H,D> and <D,H>, so the dynamics lives on a tetrahedron of population proportions. If <H,D>and <D,H> are extinct, players ignore their roles and we are back in the previous case with a unique polymorphic equilibrium. But when we consider the whole tetrahedron, this equilibrium becomes unstable, If there are a few of the other types <H,D> and <D,H> , in the population and their proportions are not exactly equal, evolutionary dynamics will move away from this equilibrium. There are lots of new equilibria created by the introduction of roles, but many are also unstable. The bottom line is that now almost every population state is carried by evolution to either All <H,D> (Maynard Smith's "Bourgeois" strategy) or all <H,D> (Maynard Smith's "Paradoxical" strategy).

To describe the same thing in different terms, Nature sends a signal to individuals, the signals are (anti)correlated, and individuals have strategies that are conditional on the signal. In the population states All <H,D> or in All <D,H> we have realizations of a special case of what Aumann (1974) calls a *correlated equilibrium*. Now let us bring the discussion back to context dependence. For situations where the role is unclear – an ambiguous signal or no signal at all – we should expect a polymorphism of Hawks and Doves. For situations with a clear signal, we should expect a correlated equilibrium.

The situation of Hawk-Dove is very similar to the symmetry observed in bargaining games. In the Nash bargaining game there are both symmetric and asymmetric equilibria. There are two symmetric equilibria, one is the fair equilibrium where each person demands ½ of the good and one is the greedy equilibrium where both demand all of the good. There are also infinitely many asymmetric equilibria where one person demands x% of the good and the other demands 100-x% of the good.

Restricting our attention to a game with just three strategies: demand 1/3, demand ½, demand 2/3, we find there are three equilibria. One where the first person demands 1/3 and the second demands 2/3, one where the first demands 2/3 and the second demands 1/3, and one where both demand half. If we consider the one population replicator dynamics again, we find that most populations go to the symmetric equilibrium but a not insignificant number go to an inefficient mixed equilibrium where the population is made up of both 1/3-types and 2/3-types (Skyrms 1996; see illustration 3). This mixed state is inefficient, like in the Hawk-Dove case, because sometimes 2/3-types meet themselves and get nothing and sometimes 1/3-types meet themselves and leave some of the good unused.

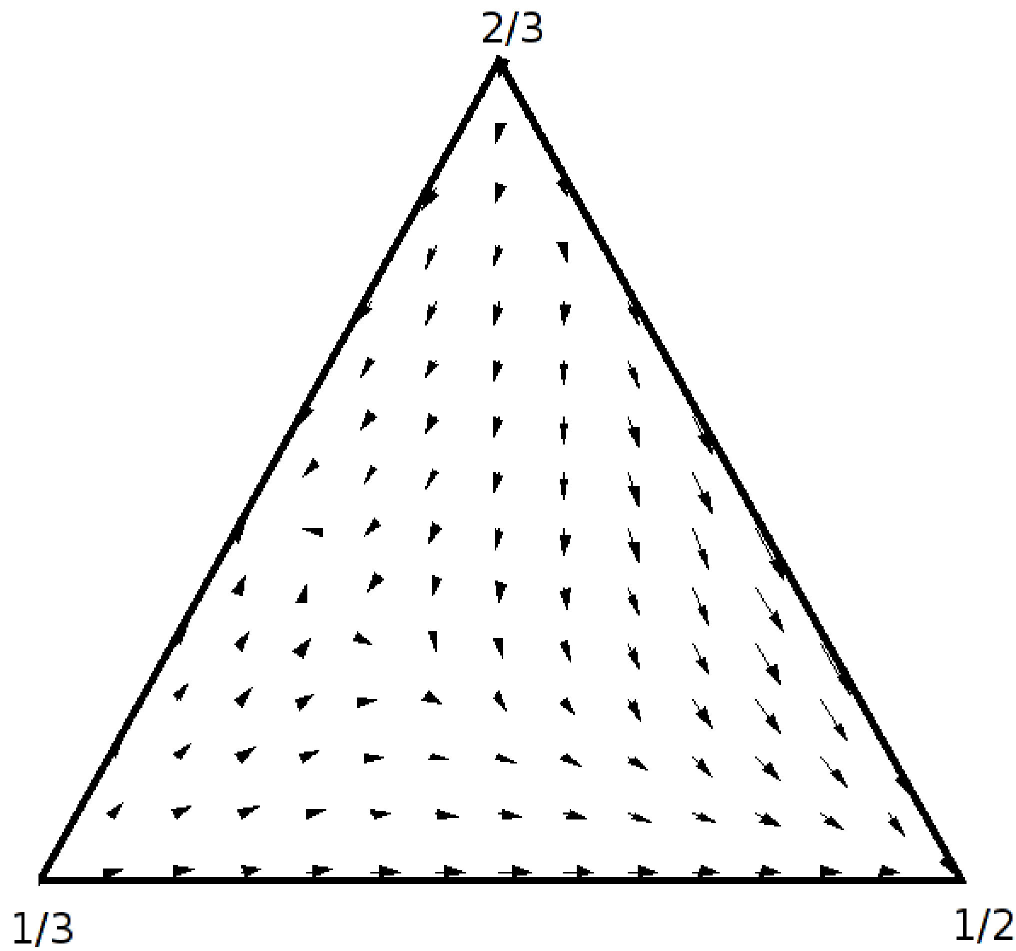

Illustration 2: Evolutionary dynamics of the Nash bargaining game

Suppose, however, that players receiver a signal – blue or green. Then there are available new strategies <1/3 if B, 2/3 if G>, and so forth. Suppose that nature assigns roles with a coin flip, and matches blues with greens. Then, just as before, the polymorphism between Always 1/3 and Always 2/3 types becomes destabilized by the addition of these new role-based strategies. And we now have new stable equilibria where the whole population plays <1/3 if B, 2/3 if G> and where the whole population plays <2/3 if B, 1/3 if G>. We also have a stable equilibrium where the whole population demands ½, regardless of signal.

One respect in which cultural evolution differs from biological evolution is with regard to the method of transmission. Although each individual in our bargaining game in endowed with a contingency plan for each circumstance (what to do if blue and what to do if green), this need not be transmitted in whole to the next "generation" of strategies. Individuals might imitate one person's contingency plan for blue and another person's contingency plan for green if they are differentially successful in the different contexts.

This possibility is modeled by a slight variant of the replicator dynamics discussed above. Instead of having a single population with full contingency plans evolving, we can treat the different contingency plans as evolving separately from one another. If blues always play against greens, this results in a model of evolution known as the two population replicator dynamics. What-to-do-if-blue strategies compete against other what-to-do-if-blue strategies, and similarly for contingency plans for green.

In this model we find that asymmetric equilibria become more likely. Considering a game with only three proposals (1/3, ½, and 2/3), almost half of the initial starting points evolving to situations where one type demands 2/3 and the other demands 1/3. The remaining populations evolve to a state where both demand ½.

We now have interesting possibilities for framing depending on whether signals are received or not. That is to say that we can evolve one norm in the absence of signals and other norms in the presence of signals. The asymmetries between social classes come to mind where division may be egalitarian within classes but inegalitarian between classes. Since strategically irrelevant properties can break the symmetry, we may see how the process of formation of social classes may originate. (For a similar observation in a spatial model see Axtell, Epstein and Young.)

Signals can indicate the structure of the underlying game, as in our first example of Nash and Ultimatum bargaining. No-cue behavior may be different from cued behavior. Signals may break symmetry and enable correlated equilibria in a single underlying game, as in our other examples. No-cue behavior again may be different from cued behavior. More generally, the point can be generalized to more complex systems of signals for more complex classes of games, which may combine these effects.

There is nothing very surprising in our general point, and we think that most experimentalists would agree. For instance, in a review of the literature on repeated Stag Hunt games, Von Huyck, Battalio and Rankin (1997) find that on average, subjects walk into the laboratory with an inclination towards the payoff dominant solution (Stag Hunting), but learn to follow risk dominance (Hare Hunting) as the experiment is repeated. Obviously, the norms that subjects initially use are not formed by repetitions of the same interaction used in the experiment, but some broader class of interactions.

In an intriguing follow-up experiment by Rankin,Von Huyck, and Battalio (2000), subjects played a sequence of different Stag Hunt games without labels, so that the only information they have is payoff information. Three out of four groups coordinated on a norm of following the payoff dominant equilibrium and the fourth group appeared on the way to converging to this norm. This unique experiment points the way to more general type of experiment relevant to framing of norms.

On a larger scale, some norms may be formed by classes of interactions so broad as to be thought of as constituting a culture. We take this as the moral of the multi-cultural studies in Henrich et. al. (2006). These shatter a lot of universal images – not only that of *homo economicus*, but also alternatives formed in experiments on college students in modern developed countries. The way individuals play standard games varies dramatically from one culture to another. It is evident that other kinds of customary social interactions in the culture play a role in shaping norms that are applied public goods provision games or ultimatum bargaining. The challenge of studying the origin and framing of systems of norms has implications for experimental economics, but it also calls for empirical research outside the laboratory.

References:

Alexander, J. M. (2007) *The Structural Evolution of Morality.* Cambridge: Cambridge University Press.

Aumann, R. (1974) Subjectivity and correlation in randomized strategies. *Journal of Mathematical Economics* 1:67-96.

Axelrod, R. and Hammond, R. A. (2006) “The Evolution of Ethnocentrism” *Journal of Conflict Resolution* **50:** 926-936.

Axtell, R. Epstein, J. and Young, H. P. (1999) “The Emergence of Classes in a Multi-Agent Bargaining Model” In *Social Dynamics* ed. S. Durlauf and H. P. Young. Ocford: Oxford University Press.

Bicchieri, C. (2006) *The Grammar of Society: The Nature and Dynamics of Social Norms.* Cambridge: Cambridge University Press.

Binmore, K., Gale, J. and Samuelson, L. (1995) "Learning to be Imperfect: The Ultimatum Game" *Games and Economic Behavior* **8:** 56-90.

Güth, W., Schmittberger, R. and Schwarze, B. (1982) "An Experimental Analysis of Ultimatum Bargaining" *Journal of Economic Behavior and Organization* **3**: 367-388.

Harms, W. (1997) "Evolution and Ultimatum Bargaining" *Theory and Decision* **42:** 147-175.

Henrich, J., Boyd, R., Bowles, R., Camerer, C., Fehr, E. and Gintis, H. eds. (2004) *Foundations of Human Sociality:Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies* Londer: Oxford University Press.

Maynard Smith, John (1982) *Evolution and the Theory of Games*. Cambridge: Cambridge University Press.

Mehta, J., Starmer C. and Sugden, R. (1992) "An Experimental Investigation of Focal Points in Coordination and bargaining: Some Preliminary Results" In J. Geweke (Ed.) *Decision Making Under Risk and Uncertainty: New Models and Findings.* Norwell: Kluwer.

Mengel, F. (2008) "Learning Across Games" *Working Paper*

Rankin, F. W., Van Huyck, J. B. and Battalio, R. C. (2000) “Strategic Similarity and Emergent Conventions: Evidence from Similar Stag Hunt Games” *Games and Economic Behavior* 32: 315-337.

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

Thaler, R. H. (1988) “Anomalies: The Ultimatum Game” *Journal of Economic Perspectives* 2: 195-206.

Van Huyck, J. B., Battalio, R.C. and Rankin, F. W. (1997) “On the Origin of Convention: Evidence from Coordination Games” *Economic Journal* 107: 442:576-596.

Zollman, K. J. S. (2008) “Explaining Fairness in Complex Environments” *Philosophy, Politics and Economics* 7, 81-97.