How should a Behavioral Economist do Welfare Economics?

H Lorne Carmichael — Queen's University

W Bentley MacLeod* — University of Southern California

July 2002

Abstract

Economists use the standard rational model to predict behaviour under a new policy regime and to evaluate the policy according to its impact on the welfare of the people affected. Experimental observation of behaviour casts some doubt on the predictive accuracy of the standard model, but the more realistic behavioral alternatives often provide a poor basis for normative evaluations. This paper suggests that in some cases we can do both. A behavioral trait can be modeled as a cognitive strategy that has evolved to augment a deeper notion of personal welfare. This makes it possible to predict behaviour with greater accuracy and to make normative evaluations of the outcomes of policy.

JEL Classification: D0

Keywords: bargaining, fairness, property rights, endowment effect, framing

*The ideas in this paper were first presented at Queen's University in April of 1994 and at a conference on evolutionary modelling at Wilfred Laurier University in August of 1995. The authors would like to thank Jennifer Arlen, Jordi Brandts, Don Cox, Jack Hirschleifer, Hervé Moulin, Vai-Lam Mui, Louis Putterman, Vernon Smith, Matthew Spitzer and Dan Usher for helpful comments, many other people for encouragement, and would also like to thank the S.S.H.R.C. of Canada and the National Science Foundation, grant SBR-9709333, for financial support of this project.

A methodological divide has arisen in the theory of economic behaviour. In order to account for their experimental results, behavioral economists have begun to consider "augmented" utility functions that when maximized subject to objective constraints will predict observed behaviour. The classic example is probably the "Prospect Theory" of Kahneman and Tversky (1979), but there are many others including Rabin (1998) who argues that normative notions of fairness should be included in the utility function, Fehr and Schmidt (1997) who argue that concern over relative payoffs should be included, and Falk and Fischbacher (1998) who include concerns for equity, reciprocity, and the intentions of others.

In spite of the clear experimental data, many theorists and applied mainstream economists are continuing to use utility functions that satisfy the standard axioms of rational choice. An argument for this position is articulated by Myerson (1999), who states simply that:[1] "In order to handle normative questions there must be some concept of human welfare in the model.". Normative economics begins with the assumption that social welfare is determined by the individual welfare of members of society. But augmented utility models are not always very useful here. For example, preferences that incorporate "loss aversion" Kahneman and Tversky (1979) do not even satisfy transitivity[2] and cannot, therefore, provide a meaningful basis for making personal welfare judgements.[3]

The problem for policy oriented behavioral economics is a serious one. In order to evaluate an economic institution we must have a model of preferences that can serve as a basis for meaningful welfare judgements. But the model must also be able to predict behaviour in the presence of the institution under study. Otherwise the results and any

[1] Myerson (1999), p. 1069

[2] See Loomes, Starmer and Sudgen (1991) .

[3] A different issue arises arise when there is an "interaction" term in agent's utility functions. These interactions are external to the market, but the indicated policy interventions are often unpalatable. A behavioral utility function gives no guidance as to which of its elements (all of which might affect behavour) should carry moral force in a welfare calculation. These issues become enormously complex, and we do not address them here. Kaplow and Shavell (1999) provide an extensive discussion with many references. See also Sen and Williams (1982).

policy implications will be vacuous. As things stand now, neither the rational approach nor the augmented utility approach, by itself, can satisfy both requirements.

Each side in this debate will argue that on balance its approach is the better one, but it is not our purpose to favor one or the other. Rather, we hope to propose a small step toward a resolution of the basic dilemma. We hope to show, in a very particular context, that one can reconcile the two approaches.

In this paper we build a theory of the endowment effect[4] and the related divergence between observed willingness to pay (WTP) and willingness to accept (WTA). This behaviour is not simply assumed to be a part of individual preferences. Underlying all individual behaviour is a standard, well behaved utility function. We show that agents who behave *as if* their preferences are intransitive will gain an advantage in bargaining and trading situations. It follows that behavior in a trading situation does not directly reveal preferences.

Our approach is an extension of the insights of Schelling (1980) and Frank (1988), who argued that in bargaining situations a willingness to be intransigent can improve one's position – a result that is formally explored in Crawford (1982).[5] We take as given this ability to commit, and model it as a willingness to walk away from a deal if the offered amount is less than "fair". The behaviour is similar to that of animals defending a territory – in each case agents will undertake actions that are costly to both parties rather than accept an amount that is too low. The territorial boundaries we examine (i.e. "what's fair") are modelled as the outcome of an evolutionary process at the cultural level. An equilibrium configuration of territorial claims is a system of natural property rights. In such an equilibrium agents will defend what they "own" and, equally important, allow others to enjoy what they own.

Throughout the paper we use the word "fairness" in this sense of "my fair share" or "fair territory". The term is not normative. We will use other terms such as "equitable" to

[4] See Coursey, Hovis and Schulze (1987), Knetsch (1989), and Knetsch and Sinden (1984).

[5] Important recent work in this area includes Ellingsen (1997) and Huck, Kirschsteiger and Oechssler (1997), who establish the evolutionary value of commitment in bargaining and trading situations.

describe outcomes that might satisfy standard normative criteria.

In Section 1 we introduce a simple trading environment. We first examine the case where agents enter a match with an initial allocation but have no reason to trade one good for another. We show that territorial behaviour, where each agent claims and receives the right to walk away with their initial allocation, will be part of any efficient equilibrium outcome. This result is a standard demonstration of the benefits of private property, and is not new. However when we extend the model to include the possibility for exchange we find that individuals will also develop an inherent "unwillingness to trade" that is consistent with the endowment effect – all acceptable trades must make each agent strictly better off. This leads to an observed difference between willingness to pay and willingness to accept, and revealed preferences are therefore intransitive.

The behavior we model is an example of a "framing effect", in that the behavior of agents depends systematically on what appear to be nonessential aspects of their situation. Section 2 discusses the case for accepting our approach as a more general explanation for framing effects in nonstrategic environments. The mechanism for the argument is plausible, given what recent work in Cognitive Science has revealed about the structure of the brain and given the general way that framing effects manifest themselves in the laboratory. The key, however, will be the ability of the approach to predict the effects of framing on behavior in new choice situations. We review some of the available evidence and make some suggestions for further experiments.

The idea that revealed preferences and "true" preferences might not be the same thing is fairly uncommon in Economics, although of course it is not new.[6] The danger in separating true preferences from behaviour is that one loses what many economists believe to be the only trustworthy source of information about individual welfare. In our framework, however, revealed preferences are derived from true preferences in a systematic

[6] Harsanyi (1955) acknowledges that there may be a difference between "individuals' actual preferences [and] their "true" preferences, that is, the preferences they would manifest under "ideal conditions," in possession of perfect information and acting with perfect logic and care."

way. Thus one might still be able to observe behaviour and make inferences about true preferences. These issues are discussed in Section 3. We follow up with an explicit suggestion for the evaluation of public projects that may prove useful in contingent valuation studies.

It is worth emphasizing that our explanation for the intransitivity of revealed preferences is entirely economic in nature. It depends only on the assumption that humans have developed strategies to help them survive in a world of finite resources, competing wants, and the potential for costly disputes. Most important, since we are able to characterize outcomes in light of a stable underlying preference ordering, it is possible to use this theory of behaviour as a basis for making welfare judgements in situations where inconsistent revealed preferences would seem to make this impossible.

1 A Trading Model

We humans differ from even the most social of animals in that we engage in the peaceful trade of one commodity for another. But before two people can agree to trade, they must first accept each other's right to walk away with his endowment intact. Economists have long argued that an important role of government is to enforce these property rights. However, our respect for property also has aspects of a social convention. It is taught to us as children and observed in a largely unreflective manner by market participants. All of us, through our behaviour, appear to attach a much higher value to what we "own" than to what we do not own. Our behaviour is similar to that of other animals when they claim and defend a territory.

In this section we assume that trading agents can be territorial. This assumption will not be defended until later – here we simply work out its implications in a very simple trading environment. This environment differs from the standard one in that initially we do not assume the *ex ante* existence of any particular set of property rights. We show first that under certain conditions the unique efficient equilibrium involves attaching a property right to the initial endowment. A second result is that successful agents will develop an inherent "unwillingness to trade" away from their endowment point.

We do not attempt to model explicitly the process by which adults might have learned the amounts to which they are entitled in their society. Instead we make use of some results that suggest that Evolutionarily Stable (Maynard Smith (1982)) outcomes will arise under a wide variety of assumptions about the underlying dynamics of cultural change.[7] We also use the well known result that strict Nash equilibria are evolutionarily stable. This allows us to simplify the analysis enormously, and yet we can still make predictions about the character of the outcomes that will be deemed fair under different circumstances.

1.1 Property Rights

In this subsection we introduce the basic model and make a very simple point about property rights. Many writers have emphasized the efficiency enhancing aspects of a system of property rights, and many of those, including Buchanan (1975), have also remarked on the fact that our familiar system is fully decentralized. In trading with our grocer, we need know nothing about his character or his wealth. We simply accept his right to keep what he has, and he accepts our right to walk away without buying anything.

We argue here that our lack of knowledge about his wealth is an important complement to our system of property rights. With private information there is a unique efficient equilibrium where each agent will protect the last dollar in her pocket but is not much interested in obtaining the first dollar in anyone else's pocket. Absent some central authority this outcome will not obtain in a standard model where utility increases smoothly with wealth. In such a model someone who values wealth more heavily than you would work harder to take your possessions than you would to defend them.

We consider a very simple matching game in which agents meet, each having in their physical possession some amount of a commodity. Agent i has in his possession the amount $e^i > 0$. For simplicity, we assume that these amounts take on a finite number of possible

[7] Weibull (1995) provides a good discussion.

values, taken from the set $\Phi = \{e_1, e_2, ..., e_n\}$, each occurring with probability q^k. The amounts are randomly assigned to agents before each match. There are no gains from trade - when agent i meets agent j, the total surplus to the match is just: $S = e^i + e^j$.

In principle there are two factors that might influence the bargaining outcomes for an agent within a match. The first is the agent's territorial claim, which we denote by $\sigma \in [0, \infty)$. This takes the form of a commitment to start a fight if the offered amount is less than what this agent believes is fair. (This might happen, for example, if the agent enters a match with some positive amount and her partner suggests she should leave with nothing.) If the available surplus is not large enough to allow both agents at least as much as they believe is fair, then agreement will not be reached and both agents will get nothing.

The second factor is an agent's ability to bargain over any surplus that remains unclaimed after each party has established his territory. We simplify by assuming agents have the same ability to bargain over unclaimed amounts, so that any such amount is divided evenly.

The stages in the history of a match between individuals i and j can be summarized as follows.

1. Agents are born at the beginning of a period and each agent is assigned an endowment $e^i \in \Phi = \{e_1, e_2, ..., e_n\}$, with q^k the proportion of agents given type k.

2. The individuals i and j are matched.

3. Each agent forms an emotional commitments σ^i that may depend on his type e^i and the total surplus S if this is known

4. Agents then play a bargaining game with the following reduced form payoff:

$$u^i\left(\sigma^i, \sigma^j\right) = \begin{cases} 0, \text{ if } e^i + e^j - (\sigma^i + \sigma^j) < 0 \\ \sigma^i + (e^i + e^j - (\sigma^i + \sigma^j))/2, \text{ if not.} \end{cases} \tag{1}$$

5. Agents die, and are replaced with a new generation of agents next period[8].

[8] The assumption that agents live for one period does not affect the equilibrium outcomes. It avoids potential issues of history dependence in payoffs.

We assume that each agent plays a pure strategy, and we define a population strategy as a mapping $\widetilde{\sigma} : \Theta \to [0, \infty)^n$ that assigns a strategy to each type in the population. Individual agents do not play mixed strategies, therefore, but there may be a (finite) mix of strategies present in the population. Let Σ be the set of population strategies for this game and consider the strategy $\widetilde{\sigma} = (\sigma^1, \ ..., \ \sigma^n) \in \Sigma$. The ex ante expected payoff to an agent from this population when matched with an agent from a population playing strategy $\widetilde{\sigma}'$ is given by

$$u(\widetilde{\sigma}, \widetilde{\sigma}') = \sum_{k,l \in \{1,...,n\}} u(\sigma^k, \sigma^l) q^k q^l \tag{2}$$

Consider first the case where the initial allocations of the agents are public knowledge. One possibility is for each agent to demand the amount $\sigma = (e^i + e^j)/2$. This set of demands will clearly form a strict Nash equilibrium, and note that there will never be any disagreements.

A pure sharing society, where each member feels he has a right to an equal share of the resources available, is therefore evolutionarily stable and efficient so long as there is full information about the initial allocation of each agent. Many other equilibria are possible as well, both efficient and inefficient.[9]

Suppose now that the initial allocations are private information. In this case the total surplus is not known before individuals enter into bargaining, and we have the following proposition:

Proposition 1 *When the initial allocation of each agent is private information, the unique efficient evolutionarily stable strategy is* $\sigma^e = e$ *for every* $e \in \Phi$.

Proof. Suppose all agents demand $\sigma\left(e^k\right) = e^k$. Then there is always agreement, and the

[9] Ellingsen (1997) studies a similar model. There are some other results that attach special status to the 50/50 split. Young (1993) has shown in an evolutionary model of bargaining that the equal division rule is stochastically stable. Given that the initial allocations correspond to "sunk investments", then Dawid and Macleod (forthcoming) show the equal division rule continues to be stable so long as endowments are known.

outcome is efficient with an expected return given by:

$$u^* = \sum_{l=1}^{n} e^k q^k.$$

It is straightforward to verify that this is an ESS.

It is also the unique efficient ESS. First observe that if for some endowment a positive fraction of agents play $\sigma\left(e^k\right) > e^k$ then there would be disagreement in equilibrium (when type e^k plays e^k), and hence this would be an inefficient outcome. Now suppose that in equilibrium some strategy has $\sigma\left(e^k\right) < e^k$. But in this case there must be another type l with $\sigma(e^k) + \sigma\left(e^l\right) = e^k + e^l$, or else agent k could increase her demand without causing a disagreement with any other agent, and this would increase her overall return. But now we have shown that $\sigma(e^l) > e^l$, which is impossible.

■

Whenever there is the threat of violence, the fact that something is in your possession does not guarantee that you will get to keep it. However, we have shown that an equilibrium exists where $\sigma^i = e^i$. This is a society where the amount initially in a person's possession is treated as an *endowment*. Each person in a match demands the right to walk away with what they had before the match began. Most important, each also grants to others this same right. This is the starting point for most economic analyses of trading behavior, and in a standard model an important role for government is to enforce these property rights.

We have shown here that property rights can arise without government when people have the ability to make territorial demands. We do not wish to argue that there is no role for government – clearly an essential part of our model is that both sides can force the bad outcome. When one party is armed with a weapon the situation can be different, and there is a clear role for government to enforce its monopoly on the use of force and prevent a personal arms race from developing among citizens. The law may also be an important factor in selecting the efficient equilibrium, since there are many inefficient ones. Our point here is simply that the vast majority of our economic interactions do take place out of the easy reach of the police, and they seem to work out just fine. This occurs because people appear to place a much higher value on the last dollar in their wallet than the first dollar in their neighbor's wallet. This allocation of property is consistent with custom and with the law of the land, but enforcement is decentralized.

In a more primitive society where possessions cannot be hidden sharing equilibria can also be efficient. Of course the private property institution is also efficient under these conditions, although it is not unique. What this analysis suggests is that sharing societies, to be successful, must also be societies with very little privacy.[10]

The agents in this model, regardless of the particular distribution of territory in their society, behave as if their preferences satisfy "loss aversion". They behave as if they care more about losses than gains relative to their fair (i.e., territorial) demand. But this behavior does not directly reveal their preferences – it is part of an equilibrium strategy. It is the strategic nature of this behavior that allows us to predict its character, and it raises the question as to whether one might still be able to use behavioral data to recover underlying preferences. We return to this in Section 3.

1.2 Trade

In this section we extend our matching game to include the possibility that the two parties to a match will want to trade. Each agent i now enters the match with a strictly positive randomly drawn initial allocation of each of two goods, denoted by $e^i = e^i_1, e^i_2$. These amounts are private knowledge – i.e. while an agent might know the amounts his partner has offered to trade, he will not know how much he has left over. The agent also has a preference ordering over the two goods represented by $U\left(x^i_1, x^i_2\right)$, where $\left(x^i_1, x^i_2\right)$ denotes consumption levels and $U(0,0) = 0$.

This utility function is stable, transitive and increasing in both its arguments, and we assume it is a meaningful personal economic welfare ordering in the usual sense. In

[10] Yellen (1990) provides a fascinating history of a Kalahari desert society that moved from a sharing to a private property regime soon after the introduction of new stores of value (from outside contact) that could be concealed.

principle its form would differ across agents and also be private knowledge, but we do not need this for the results that follow. The assumption that endowment levels are private information is sufficient to ensure that on entering a particular match the agent cannot predict which of the two goods he will want to buy or sell, or how much. We denote the lower contour set through the initial allocation point by $L(e^i)$ with associated border denoted $\overline{L}(e^i)$.

The general situation for one of the agents is illustrated in the diagram below.

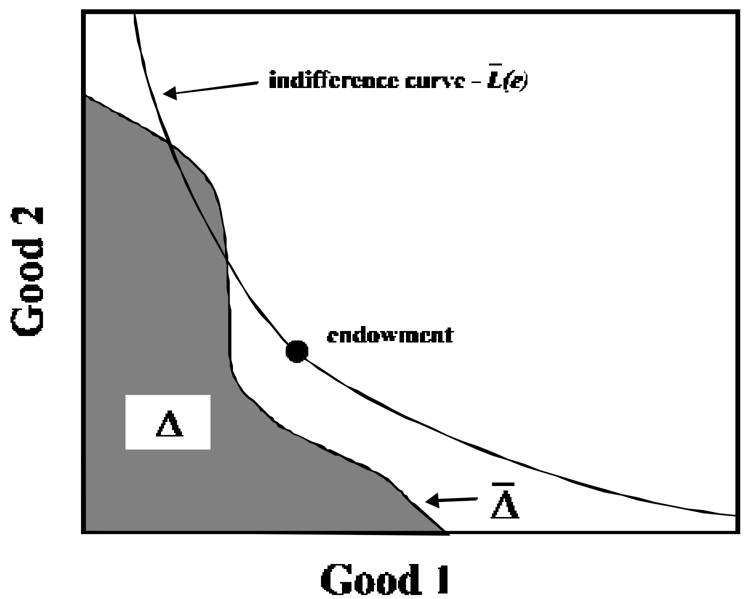

Figure 1:

The region Δ is the territory claimed by the agent whose origin is at the lower left. This region defines the agent's territorial bargaining strategy. The agent considers the outcomes in the interior of this region to be "unfair" and he will reject them even though the alternative is to get nothing, which is a worse outcome according to the agent's own preferences. Such a rejection will be "violent" in that it will also force the other party to get nothing.

To keep things as simple as possible we will assume that each agent will treat his initial allocation as his private property, and will grant this benefit to all others that he meets.[11] This is the case in all experimental investigations of the endowment effect, and also seems to describe the world in which we live. What we shall find is that in a world where people can trade one good for another, the institution of private property is no longer sufficient for full efficiency.

Bargaining outcomes are determined according to the territory claimed by each agent. Agent i will claim a two dimensional territory Δ^i, the outer boundary of which is denoted by $\overline{\Delta}^i$. For expositional simplicity we will say that an agent "claims" the boundary of his territory, although it is only those points interior to the set that lead to disagreement. We assume this boundary is continuous and for consistency we also assume that if a point $x = (x_1, x_2)$ is claimed, then so are all points y such that $y \leq x$.

Note that in communicating with each other the agents will have to speak in terms of net trades from their initial allocation points, since their endowment levels are private. In the text that follows we will continue to speak in terms of absolute consumption and allocation levels.

When two agents i and j meet there may be a region in the Edgeworth Box containing allocations that both parties would find acceptable. Denote this region by Υ^{ij}, and note that this region is a closed set. By construction all elements $(x^i, x^j) \in \Upsilon^{ij}$ are physically feasible, in that the allocation to each agent adds up to the total allocation to the match. We assume that the two agents will dicker agreeably within the unclaimed region if it is not empty. Since each party respects the private property regime we have $(e^i, e^j) \in \Upsilon^{ij}$. Since the initial endowment is always available we will assume that it also forms the the default payoffs in any bargain reached by the two individuals.

We assume the bargaining outcome for agents i and j are given by the Nash bargaining solution:

$$\arg \max_{(x^i, x^j) \in \Upsilon^{ij}.} \left(U^i \left(x^i \right) - U \left(e^i \right) \right) \left(U^j \left(x^j \right) - u \left(e^j \right) \right) \tag{3}$$

[11] We believe that a result similar to Proposition 1 is available for the two dimensional case, but a proof has so far eluded us. The bargaining set in the two dimensional case with general territorial demands becomes very unstructured.

where we note that by construction (x^i, x^j) is physically feasible. This solution will pick a point on the contract curve that makes each agent better off than she would be without trade.

We wish to characterize the strategies in an evolutionarily stable population, where a strategy is simply a claimed region Δ. We note first that it is a Nash equilibrium for each agent to claim the "no trade" region $\Delta^* = \{x \mid x_1 \leq e_1 \bigvee x_2 \leq e_2\}$. In this case each agent claims her initial allocation point, and will only accept those trades that provide more of both goods. The only feasible outcome is the initial allocation. This is a Nash equilibrium since claiming more of either good will cause disagreement, and claiming less cannot make you any better off. In equilibrium there will be no trade, but there will be no violence and each agent will be able to leave with what she brought to the table.

The no trade outcome is a Nash equilibrium but it is not evolutionarily stable. To be evolutionarily stable a population must be immune to invasion by a small group (rather than just an individual) playing a different strategy. Suppose a small group entered made up of agents who claim $\Delta = L(e)$. These agents will not be able to trade with incumbents (nobody can), but would trade when they met each other. Thus they would do better than incumbents in the new mixed population, and the no trade population is not evolutionarily stable.[12]

The most efficient outcome, of course, would be achieved by a population where each agent claimed $L(e^i)$ and no more. It follows easily from the geometry of the Edgeworth Box that there would always be agreement, and whenever there are gains from trade these would be fully realized. This is the standard textbook case, where it is assumed that people will trade whenever the gains from trade are positive. However, an agent may be able to gain if he is willing to risk leaving some gains from trade on the table.

[12] It is also easy to see that the no trade outcome is not a strict Nash equilibrium

Suppose a new agent enters a population where all agents are claiming $L(e)$ with a claimed territory as in Figure 2.

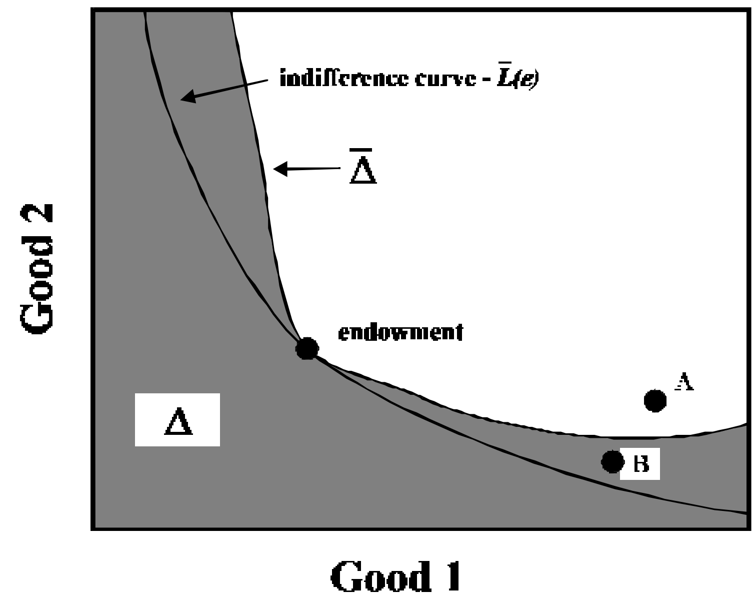

Figure 2:

This agent's territorial border includes his endowment point, but everywhere else $\bar{L}(e)$ is strictly contained within Δ. Since this agent does not claim any points that have more of both goods than his endowment he is respecting the property rights of everyone in the population. The worst he and they will ever do is walk away with their endowments. But suppose he is offered an outcome like point B. This outcome increases his welfare, but this agent will nonetheless refuse such a deal, choosing instead to keep his endowment. He would accept an outcome like point A.

To see why this agent will do well, consider what happens with a very small expansion of his claimed territory. This expansion, so long as it does not involve claiming more of both goods, will initially have no effect on his trades with anyone. It will reduce the size of the bargaining set Υ^{ij} with any partner j, but it will not impinge on the solution point for the bargain (since this will be at an interior point of Υ^{ij}), nor will it affect the threat point for the Nash bargaining solution (since this is given by the no trade option). Hence the only impact would occur after the expansion of territory reaches a point where it directly affects the solution to a particular bargain. At this point a further small increase in claimed territory must increase this agent's payoff. Therefore at any evolutionarily stable equilibrium where property rights are respected some agents must be claiming strictly more than $L(e)$.[13]

Formal examination of a model with general territorial claims becomes quite complex, and evolutionarily stable populations may not exist since an individual may be able to perturb part of her set of demands in an inessential way. In the appendix we introduce a two dimensional parameterization of the possible territorial demands where all perturbations are "essential". One dimension d corresponds to how much below or above the endowment point to demand (where $d = 1$ means the endowment is claimed as private property) while the second dimension corresponds to the amount of curvature on the boundary of the claimed territory (where $\alpha = 0$ corresponds to $\overline{L}(e)$ and $\alpha = 1$ corresponds to Leontief claims - the no trade case). Increasing α above zero increases the likelihood that efficient trade does not occur. In the appendix we prove the following proposition:

Proposition 2 *In any evolutionarily stable population that respects property rights ($d = 1$), individuals demand territories with boundaries that include their initial allocation, but elsewhere strictly contain the associated lower contour sets (i.e $\alpha \in (0, 1)$)*[14].

We have not been able to prove that evolutionarily stable populations exist under more general assumptions about preferences and territorial claims. However, given that endowments are respected, arguments along this line will suggest that the system will not

[13] There are other formal ways that an agent's "unwillingness to trade" might increase his return from those Nash bargains that are nonetheless completed successfully. For example, if the offered trade vector is relatively long an agent could behave as if his endowment was worth more to him than it really is. We have chosen our approach to make the clearest possible distinction between preferences and behavior (i.e. claimed territory).

[14] Even for this highly parameterized case, one can prove the existence of an evolutionarily stable population only for special cases. One situation is when there are only two different endowments. In that case the payoffs are quasi-concave in α, and hence an equilibrium can be assured.

approach a situation where $\Delta^i \subseteq L(e^i)$ for any i. The dynamic will always be rewarding agents who push away from $\Delta^i = L(e^i)$ in the direction of less trade. Thus the general result that agents should be unwilling to trade for small gains seems robust.

The initial allocation plays a very special role in a trading game because it defines the dimensions of the Edgeworth Box. Even when agents know nothing about their partner's preferences or endowment they do know that they have nothing to lose and much to gain by claiming their initial allocation. This is what leads to efficiency in the one dimensional case, and it is the source of the endowment effect when there is the possibility for trade. In a private property equilibrium the initial endowment has no effect on preferences, but it has a big effect on behavior.

2 Reason and Instinct in Decisionmaking

2.1 A bicameral model of economic decisionmaking

The model discussed here is unusual in that agents get to choose something that is not normally assumed to be within their choice set – i.e. a level of commitment. Recent work in Cognitive Neuroscience provides some support for this approach. It is now known there are at least two separate systems in the human brain that take environmental cues or stimuli and transform them into actions. Higher thought and reasoning works through the cerebral cortex, but the information processing that occurs here can be quite slow and therefore not appropriate for all situations. Sometimes a stimulus gets routed by the thalamus directly to the amygdala where a more primitive level of processing occurs.

As LeDoux (1996) discusses in some detail, the "low road" to the amygdala allows individuals to respond very quickly to their environment.[15] Hence if one is driving and sees

[15] Donald (1991) also provides a very thorough and accessible survey of work on the structure of the human brain. The brain is built of many specialized parts for language, face recognition, and many other aspects of cognition, but the allocation of space seems to depend on the needs and uses established in early (and even adult) life. No physical location for the "central processor", or "seat of consciousness" has yet been

an oncoming car in her lane, she is immediately alert and may act "without thinking" to avoid a collision. The reaction is often called "instinctive", but in many cases it is clearly a response that has been learned at an earlier time - steering for the ditch has no part in our evolutionary history, and the appropriate ditch can be on the right or the left depending on the country where one is driving. The low road achieves a higher speed by using more primitive signal processing. In particular it can be activated even though after the fact one realizes it has been a false alarm. Once a response has been learned it can be difficult to unlearn.[16]

In our model of behavior an agent's welfare is being served by these two separate decisionmaking systems. One is fast acting but imprecise in that it may group several different situations into a single class that requires a single kind of response. The other is slower but more flexible and accurate. The two systems can sometimes work against each other. For example, in a famous example of reflex response Charles Darwin placed his face close to the glass of a snake display. When the snake attacked he automatically stepped back, even though he had consciously commanded his head to stay close to the glass.

The "low road" system is dominant in other animals[17], and clearly it can make mistakes. Moths have developed a method of flying straight that still works very well for them in the daytime, but now dooms them to spend their evenings circling lamposts. A grouse will burst out of the grass at the approach of a predator – a behavior which was once adaptive but now provides sport for human hunters. These animals cannot make fine distinctions among the situations they encounter, and classify them as identical.

It is our hypothesis that a similar "low road" misclassification can occur in humans, and that this is the source of "framing" effects such as loss aversion and the endowment effect. These "low road" behaviors will sometimes arise in situations where they are not appropriate. However, we humans also have the ability to reason, and over time we should

discovered.

[16] Although not impossible. People who move to a country where people drive on the other side of the road will learn to react safely.

[17] In humans this system is sometimes called the "reptilian" brain, since its structure is similar.

be able to "see through" instincts that are dysfunctional or unhelpful. This "bicameral" model of decisionmaking has testable implications for behavior, to which we will now turn.

2.2 Choice

Suppose there is a "low road" routine in the human brain that identifies trading situations and generates a territorial claim to the initial allocation and an unwillingness to trade, or "status quo" response. Further, suppose that this system is triggered when subjects in the laboratory are given objects and asked to trade for something else, even though there will be no bargaining over the terms of trade. What sort of implications would this have for observed behavior?

First of all, in the laboratory we might observe something like what is described in Figure 3.

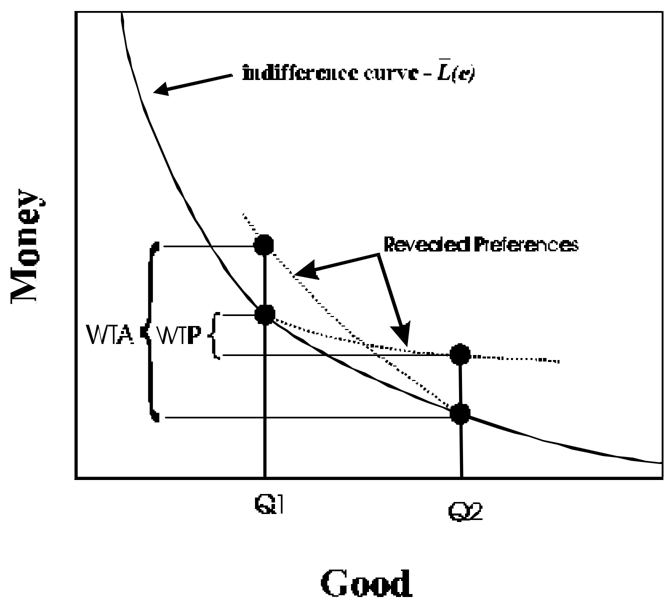

Figure 3: Willingness to Pay (WTP) versus Willingness to Accept (WTA)

Suppose someone is given a quantity $Q1$, then asked how much she would be willing to pay to move to $Q2$. The answer she would give would be less than the amount required to make her give up $Q2 - Q1$. The willingness to pay for this move is less than the willingness to accept for a move in the other direction.[18] Moreover, a true valuation of the change from $Q1$ to $Q2$ exists and it lies between WTP and WTA. As we shall see below there may be direct ways to reveal it.

Note that a researcher who believed that behaviour in simple nonstrategic choice situations reveals underlying preferences would be forced by the behaviour in this experiment to conclude that the subject has preferences that are intransitive. In fact his experiment does not reveal underlying preferences at all. Behavior in this case is revealing a strategy that happens to be inappropriate for the situation.[19]

What kind of evidence would further support the bicameral decisionmaking model? First, the way in which status quo effects arise should reflect generally what we know about behaviors that are mediated through the direct route to the amygdala. In short, they should be fast acting, and sometimes difficult even for the people involved to explain. Also, as individuals get more experience with the decision problem, and learn how to think it through, the effect in *non-strategic* situations should diminish. This seems to be consistent with the evidence. For example Francoise, Kujal, Michelitsch, Smith and Gang (1996) show that with experience the difference between WTP and WTA diminishes, though it never completely disappears.[20] It would be interesting to see if the endowment effect gains strength when these experienced individuals are placed in situations that clearly involve bargaining and trading. Another piece of evidence comes from Loewenstein and Adler (1995), who find that the endowment effect is very fast acting, in the sense that individuals change their valuation of a good immediately after being given possession. Even more interesting is the fact that individuals' predictions of how they would behave if they had a

[18] Loomes and Sudgen (1983) have shown that the difference between WTP and WTA is consistent with the endowment effect. (See Thaler (1980) for a review).

[19] Chess is a game that uses "high road" reasoning. When we see a chessplayer sacrifice his queen we do not assume he does not like this piece, or that he wants to lose. The behavior makes sense in the strategic context in which it is made, which ultimately is internal to the agent. It could be a bad move.

[20] Erev and Roth (1996) present a great deal of evidence that invididuals alter their behavior with experience, and hence could not have choosen optimally the first time they are exposed to a problem. Initial behavior is systematic, which is what we focus on. Smith (1994) reports similar results in several contexts.

good were different than their actual behavior once they had it. Their conscious brains systematically underpredicted the strength of the status quo effect they would be feeling only moments later.

A further direct test of our model would investigate the relationship between privacy and property rights. For example, subjects could be matched and allowed to trade one good for another. Failure to reach an agreement about how much to trade (if anything) would lead to each person getting nothing. In one treatment the initial allocation could be private knowledge and in another this allocation could be told to each subject. We predict that in the private knowledge case agents will respect each other's right to retain his endowment. In the public case this might not happen. Someone with a very low initial allocation, knowing that his partner has more than him, might demand a larger share of both goods backed by the threat of enforcing the zero outcome.

A second way to evaluate our hypothesis is to test the predictions it makes about behaviour in situations that are outside the context that the model was designed to explain. The following idea is illustrative. A key assumption in our model of bargaining is that the bargaining surplus is divisible. If it were not divisible then the only equitable outcome would be the one where each side gets nothing. But this is less efficient than the other possibilities where one of the agents gets to keep the good. In the absence of a "fair" mechanism for determining who wins (such as a coin toss) there really is no role for fairness to play, and the situation will just be a pure "fight" or contest.[21]

Consider now the ultimatum game. Our explanation for the observation that some responders will turn down offers of less than half the pie is that they have framed the game as a bargain over a divisible good. Subject have faced this problem many times already in a variety of different contexts, and they have developed a territorial notion of fairness that centers on the 50/50 split. They bring this cultural notion with them to the experiment

[21] This may help explain why standard auction theory predicts behavior so well. In an auction the bidders have no way to divide the gains, and none has an action that will enforce a bad outcome on everyone if this does not happen. Behavior is determined by "high road" reasoning, which standard economic theory handles well.

and responders are therefore offended if they are offered less than half the surplus. Proposers, on their part, think it inappropriate to ask for more. In practice this tendency must compete with the more rational decisionmaking that works through the cerebral cortex, and we find that neither the fairness model nor the rational model is exactly right – the proportion of offers accepted declines as the offered amount shrinks. More than half the offers of less than 20% of the pie are rejected.[22]

Now suppose we were to run an ultimatum game experiment where the proposer can offer an 80/20 split or a 20/80 split, but nothing else. In particular, 50/50 is not available. The responder can force 0/0 as usual. We predict that players will no longer frame their situation in the same way. More proposers will offer 20% and more responders will accept 20%. Neither will consider this behavior to be "unfair" under the circumstances.

Behavioral models can generate the same prediction by assuming that agents care about the intentions of others. Falk and Fischbacher (1998), for example, build such a model and present some supportive evidence. There is a great difference, however, in the welfare implications of these two approaches. We turn to this in the next section.

3 Welfare

There is no doubt that behavioral theories accomplish what they are designed to do, which is enhance our ability to predict behavior. They are less satisfactory as a basis for normative decision making because the assignment of payoffs to outcomes depends on the frame within which choices are being made. When this frame is itself based on elements from the field of choice there is simply no consistent notion of personal welfare in the model to begin with.

Given the widespread importance of framing effects on behavior it may be tempting to abandon all hope of making meaningful welfare judgements. However, if framing effects are

[22] Our approach can be distinguished from myopic learning models (see for example Samuelson (2002)) in that behavior will be anomalous at the very beginning. People have already learned how to play the game that they believe they are playing. Indeed, given the outcomes, perhaps this is the game they are playing.

themselves the outcome of an optimizing strategy then this would be premature. If something stable and meaningful really is being optimized then the only question is whether we can recover its characteristics from observations of behaviour

The following suggestion provides a further test of our decision model and may also have some practical implications for cost benefit analyses. The problem with contingent valuation experiments, according to our approach, is that subjects are in a simple choice environment but are using mental decision making procedures that were designed and are appropriate for bargaining situations. Perhaps the experimental situation could be framed in such a way as to avoid these channels. For example, the best known experimental evidence for the endowment effect comes from a series of experiments where subjects were given either a coffee mug or a chocolate bar and then given the opportunity to trade one for the other (Kahneman, Knetch and Thaler (1990)). Far less trade was observed than the standard model predicts, even when positive incentives were applied.

Suppose now that our goal as an experimenter is to find out whether a subject "really" prefers a chocolate bar to a coffee mug or a coffee mug to a chocolate bar, and we want to base this judgement entirely on the observed behavior of the subject. If our approach is correct we will never get useful information by giving the subject one of these objects and seeing if they decide to trade for the other. Rather, we might just put both objects on a table in front of her and see which one she picks up. When the decision is presented as a *free choice*, rather than a *trade*, there seems to be no reason why the subject's behavior should not reveal underlying welfare. If one of the goods is money, then after a series of trials one could determine an individual monetary valuation for the other good.

This suggests that a modification of the Equivalent Gain or the Equivalent Loss measures of welfare change should be more reliable that either Willingness to Pay or Willingness to Accept. To measure the Equivalent Gain the subject in Figure 3 at Q_1 is asked to choose between a move to Q_2 or a given amount of money. Rather than being asked whether he would be willing to pay \$500 for a larger park, for example, he is asked whether he would prefer a larger park or \$500 cash. Our model suggests that this information will be meaningful, but note that it is information about a closely related choice, rather than the one that might actually be made.

In the increasingly relevant case of government budget surpluses, it may be possible to gather good information about exactly the right choice. One can to ask people whether they would rather have $500 as a tax cut or take the park improvements instead. Obviously more work needs to be done to explore the robustness of techniques like these. The evidence in Bateman, Munro, Rhodes, Starmer and Sugden (1997), while not a direct application, suggests the approach may be promising.

4 Conclusions

Much of what we say about framing effects in this paper is not new. In developing prospect theory, Kahneman and Tversky (1979) highlight the importance of heuristics and biases for understanding human decision making. It is also understood that one of the reasons framing effects occur is that it may be faster and more efficient for individuals to use previous experiences to decide how to choose quickly, rather than to carry out a detailed analysis of the situation[23]. Similarity theory, due to Rubinstein (1988) and Leland (1994), makes this insight explicit. This theory incorporates a judgement mechanism where individuals place problems into particular categories to which they assign the same or similar payoffs. Leland (1997) presents some experimental tests.

Kahneman and Tversky (1979), Loomes and Sudgen (1983), Leland (1997) and many others approach the problem of explaining behavior by suggesting the form preferences must take in order to "rationalize" it. These theories are designed to provide a unified and parsimonious framework within which one can organize evidence. However they cannot help with normative economics since behavior depends upon variables that may not reflect the true underlying preferences of individuals, and indeed the existence of such preferences

[23] See the work by Johnson-Laird (1983) and the review of cognitive psychology by Churchland and Sejnowski (1993). See also Heiner (1983) who argues that some anomalous behavior may arise because it is adaptive in a related context.

is not always postulated.

In this paper we started with the very old fashioned assumption that at the heart of economic behavior is a meaningful welfare ordering. We endowed agents with a particular cognitive structure that permits them to act decisively before they have fully analyzed a situation. In this context we studied the important economic activities of bargaining and trading. We argued that territorial behavior, which we identify with natural property rights, will arise and we made some progress in suggesting the form that property rights will take in different situations. We showed that agents will refuse to trade if the perceived gains are small. In this way we have shown that loss aversion and the endowment effect could be be consistent with individual maximization of a meaningful welfare ordering.

Our approach can be distinguished from other recent work in three ways. First, by starting with the hypothesis (surely correct) that our character is the product of adaptive structures known to exist in our brains we can make sense out of framing effects. They are strategies that get used for a good economic reason. Second, the hypothesis does not just encompass some of the behaviour rationalized by other theories. By providing an explanation for framing effects it can generate predictions about their character in new situations. Finally, since our approach can account for framing effects while remaining anchored in the optimization of a stable individual welfare function, it does not require us to abandon welfare economics. We may be able to carry out consistent cost/benefit analyses in a way that avoids the difficulties caused by the distinction between willingness to pay and willingness to accept.

5 Appendix

5.1 An Explicit Example

It will be useful to express the agent's unclaimed territory, given his endowment, in terms of *net trades* away from the endowment point. The agent's unclaimed territory (i.e. those trades he finds acceptable) is denoted by T^i, and his final allocation, after a successful trade, is given by $(x_1, x_2) = (e_1 + t_1, e_2 + t_2)$, with $(t_1.t_2) \in T^i$.

To avoid measurability issues we assume that the set of possible endowments is a finite set of strictly positive allocations $E \subset \Re^2_{++}$, with the probability of an individual receiving $e \in E$ given by $P(e)$. Generically the game played each period is as follows:

1. Individuals are matched at the beginning of a period, with agent i receiving an endowment $e^i \in E$, with probability $P(e)$.

2. Agent i enters the match with an acceptable territory in *net trades,* $T^i \subset \Re^2$.

3. The set of net trade demands are realized, resulting in sets T^i and T^j.

4. If $T^i \cap -T^j = \emptyset$ then $U^i = U^j = 0$. Otherwise the agents bargain over the set of feasible trades $T^i \cap -T^j$.

We can now define the outcome of the process of bargaining once the territorial demands have been made. We will define the set of net trades that yield as good or better utility than one's endowment e as $T^*(e)$. We call this the set of *Paretian* trades.

In order to make analytical progress we restrict the formal analysis to the case where $u^i(x_1, x_2) = u(x_1, x_2) \equiv \sqrt{x_1 x_2}$ and the parametric set of demands, or acceptable outcomes, is defined using a constant elasticity of substitution (CES) utility function. Thus the strategy of individual i after she learns her endowment is a pair $\sigma = (d, \alpha) \in \Re_+ \times [0,1]$, corresponding to the demand:

$$T(\sigma, e) = \left\{ t \in \Re^2 | U_\alpha \left(\frac{e_1 + t_1}{e_1}, \frac{e_2 + t_2}{e_2} \right) \geq d \right\}, \tag{4}$$

where

$$U_\alpha(x_1, x_2) = \left\{ x_1^{-\tan(\alpha\pi/2)} + x_2^{-\tan(\alpha\pi/2)} \right\}^{-1/\tan(\alpha\pi/2)}. \tag{5}$$

The function U_α represents a CES utility function re-normalized so that as α varies from 0 to 1, one goes from a Cobb-Douglas utility function ($U_0(x_1, x_2) = \sqrt{x_1 x_2}$) that represents the agent's true preference, to a Leontief utility function ($U_1(x_1, x_2) = \min\{x_1, x_2\}$). The demands are normalized about the endowment in that the choice d corresponds to a utility demand premium when $d > 1$, and an acceptance of lower utility when $d < 1$. Given that consumption is non-negative, then $d \geq 0$.

Given a bargaining pair i and j, we assume the payoff given the endowments and strategies is given by the Nash bargaining solution over the set of feasible net trades:

$$v\left(e^i, \sigma^i, e^j, \sigma^j\right) = \begin{cases} 0, & \text{if } T^{ij} = \emptyset \\ u\left(e^i + t\left(e^i, e^j, T^{ij}\right)\right), & \text{if not.} \end{cases}, \tag{6}$$

where

$$T^{ij} = T\left(\sigma^i\left(e^i\right), e^i\right) \cap -T\left(\sigma^j\left(e^j\right), e^j\right), \text{ and} \tag{7}$$

$$t\left(e^i, e^j, T^{ij}\right) = \arg\max_{t \in T^{ij}} \left(u\left(e^i + t\right) - u\left(d \cdot e^i\right)\right)\left(u\left(e^j - t\right) - u\left(d \cdot e^j\right)\right). \tag{8}$$

Note that in the Nash bargain the threat point for each player is the lowest payoff in the set of feasible trades, and that within the feasible region the Nash bargaining outcome is determined by the agents' true preferences.

We shall examine the evolutionarily Stable outcomes of this game in pure strategies. The expected payoff to agent i with these strategies is formally given by:

$$V\left(\sigma^i, \sigma^j\right) = \sum_{(e^1, e^2) \in E^2} v\left(e^1, \sigma^i, e^2, \sigma^j\right) P\left(e^1\right) P\left(e^2\right).$$

Let $V^0 = \sum_{e \in E} \sqrt{e_x e_y} P\left(e^1\right)$ denote the expected payoff if each agent consumes her endowment. Our first observation is that this outcome can be achieved as a Nash equilibrium.

Proposition 3 *The strategy* $\sigma^0(e) = (1,1)$ *for all* $e \in E$*, consisting of demanding the only trades that are non-negative for both goods, forms a Nash equilibrium with a payoff of* V^0 *for each agent.*

Proof. Since $d = 1$ and $\alpha = 1$ then $T(\sigma^0(e^i), e^i) \cap -T(\sigma^0(e^j), e^j) = 0$ for all $e^i, e^j \in E$, therefore the equilibrium payoff is V^0. To see that this is a Nash equilibrium observe that for all $t \in -T(\sigma^0(e^j), e^j) \setminus \{0\}$ then $u(e^i + t) < u(e^i)$. Thus adjusting ones demand to include more potential net trades cannot increase ones payoff. Similarly, increasing ones utility demand results in no trade at all, and a zero payoff. Thus $\sigma^0(e) = (1,1)$ forms a Nash equilibrium.

∎

This result simply asserts that it is a Nash equilibrium to respect each other's property. If there are no gains to trade then this is also the most efficient equilibrium, as we see in the next proposition.

Proposition 4 *Suppose that the endowment bundle e satisfies* $supp\, P \subset \{\lambda e | \lambda \in \Re\}$*. Then all efficient equilibria are of the form* $\sigma(e) = (1, \alpha)$, $\alpha \in [0,1]$.

Proof. Given that $u(\lambda e) = \lambda u(e)$, then it is efficient for each agent to consume her endowment. Thus there are no gains from trade, except the need to reach an agreement. It is clear that in any equilibrium, if an agent with endowment e^r plays $d^r < 1$ then there must be another agent with endowment e^s who plays $d^s > 1$, such that when these two meet there is nothing left unclaimed. Otherwise agent r could increase her demand and do strictly better against s and no worse against anyone else. But agent s will disagree with itself, and the equilibrium cannot be efficient. Thus $d = 1$ is the unique efficient equilibrium, while the value of α is irrelevant.

∎

Given that the value of α is not uniquely determined, these equilibria are not ESS. But in this case there are really no gains from trade, and hence the two dimensional setup is not of much interest.

Consider now the case where endowments may not be in the same proportion, and hence there maybe gains from trade. As in the previous case there exist strategies that ensure an efficient outcome is achieved. If agents demand the Paretian net trades, T^*, this results in the most efficient payoff possible, defined by

$$V^* = \sum_{(e^1,e^2)\in E^2} u\left(e^1,\sigma^*,e^2,\sigma^*\right) P\left(e^1\right) P\left(e^2\right) \tag{9}$$

where $\sigma^*(e) = (1,0)$, for all $e \in E$.

We can show that these efficient strategies do not form an ESS. However, although the payoffs are continuous in α, they are not quasi-concave and hence we cannot in general prove the existence of any ESS. Here we will simply show that at any ESS where private property is respected, revealed preferences are not transitive. When equilibria do not exist the underlying dynamics are likely to produce some form cycling behavior.

Proposition 5 *At every ESS that respects endowments ($d^e = 1, \forall e \in E$) agents demand trades strictly inside the set of Pareto improving net trades $T^*(e)$, ($\alpha^e \in (0,1), \forall e \in E$).*

Proof. First observe that if $\alpha^e = 1$ then for this type only $t = 0$ is feasible. Hence decreasing α^e can only make this type better off, and therefore $\alpha^e = 1$ cannot be part of an ESS. Now consider the case of $\alpha^e = 0$. Increasing α^e does not decrease the probability of an agreement, nor does it affect the threat point for the Nash bargaining solution. Hence the only impact would be on those net trades that lie on type e's indifference curve. Given the continuity of payoffs in α, then this implies that a small increase in α^e must increase type e's payoff. Therefore at an ESS we must have $\alpha^e \in (0,1)$ for every $e \in E$. ■

References

Bateman, Ian, Alistair Munro, Bruce Rhodes, Chris Starmer, and Robert Sugden, "A Test of the Theory of Reference-Dependent Preferences," *Quarterly Journal of Economics*, May 1997, *112* (2), 479–506.

Buchanan, James, *The Limits of Liberty*, Chicago, IL: Chicago University Press, 1975.

Churchland, Patricia S. and Terrence J. Sejnowski, *The Computational Brain*, Cambridge, MA: MIT Press, 1993.

Coursey, Don L., John L. Hovis, and William D. Schulze, "The Disparity Between the Willingness to Accept and the Willingness to Pay Measures of Value," *Quarterly Journal of Economics*, August 1987, *102*, 679–90.

Crawford, Vincent P., "A Theory of Disagreement in Bargaining," *Econometrica*, May 1982, *50* (3), 607–637.

Dawid, Herbert and W. Bentlet Macleod, "Hold Up and the Evolution of Bargaining Conventions," *European Journal of Economic and Social Systems*, forthcoming.

Donald, Merlin, *Origins of the Modern Mind*, Cambridge, MA: Harvard University Press, 1991.

Ellingsen, Tore, "The Evolution of Bargaining Behavior," *Quarterly Journal of Economics*, May 1997, *112* (2), 581–602.

Erev, Ido and Alvin E. Roth, "On the Need for Low Rationality, Cognitive Game Theory: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria," May 1996. http://www.pitt.edu/ alroth/alroth.html.

Falk, Armin and Urs Fischbacher, "A Theory of Reciprocity," Technical Report, University of Zurich 1998.

Fehr, Ernst and Klaus Schmidt, "A Theory of Fairness, Competition, and Cooperation," *mimeo, IERE Zurich*, 1997.

Francoise, Robert, Praveen Kujal, Roland Michelitsch, Vernon Smith, and Deng Gang, "Experimental Tests of the Endowment Effect," *Journal of Economic Behavior and Organization*, August 1996, *30* (2), 213–26.

Frank, Robert H., *Passions within Reason*, New York, NY, U.S.A.: W. W. Norton & and Company, 1988.

Harsanyi, John C., "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility," *Journal of Political Economy*, 1955, *63*, 309 – 321.

Heiner, Ronald, "The Origin of Predictable Behavior," *American Economic Review*, 1983, *73*, 560–595.

Huck, Steffan, Geog Kirschsteiger, and Jörg Oechssler, "Learning to Like What You Have - Explaining the Endowment Effect," January 1997. mimeo, Humboldt University, Berlin.

Johnson-Laird, Philip N., *Mental Models*, Cambridge, MA: Harvard University Press, 1983.

Kahneman, Daniel and Amos Tversky, "Prospect Theory: An Analysis of Decisions Under Risk," *Econometrica*, 1979, *47*, 262–91.

___, Jack Knetch, and Richard Thaler, "Experimental Tests of the Endowment Effect and the Coase Theorem," *Journal of Political Economy*, 1990, *98*, 1325–48.

Kaplow, Louis and Steven Shavell, "Principles of Fairness versus Human Welfare: On the Evaluation of Legal Policy," Technical Report 1999. manuscript, Harvard University Law School.

Knetsch, Jack L., "The Endowment Effect and Evidence of Nonreversible Indifference Curves," *American Economic Review*, 1989, *79*, 1277–84.

—— **and John A. Sinden**, "Willingness to Pay and Compensation Demanded: Experimental Evidence of an Unexpected Disparity in Measures of Value," *Quarterly Journal of Economics*, August 1984, *99*, 507–21.

LeDoux, Joseph, *The Emotional Brain*, New York: Simon and Schuster, 1996.

Leland, Jonathan W., "Generalized Similarity Judgements: An Alternative Explanation for Choice Anomalies," *Journal of Risk and Uncertainty*, 1994, *9*, 151–172.

——, "Similarity Judgements in Choice Under Uncertainty," *forthcoming Management Science*, 1997.

Loewenstein, George F. and Daniel Adler, "A Bias in the Prediction of Tastes," *Economic Journal*, July 1995, *105*, 929–937.

Loomes, Graham and Robert Sudgen, "A Rationale for Preference Reversal," *American Economic Review*, June 1983, *73* (3), 428–432.

——**, Chris Starmer, and Robert Sudgen**, "Observing Violations of Transitivity by Experimental Methods," *Econometrica*, March 1991, *59* (2), 425–439.

Myerson, Roger B., "Nash Equilibrium and the History of Economic Theory," *Journal of Economic Literature*, September 1999, *37* (3), 1067–1082.

Rabin, Matthew, "Pyschology and Economics," *Journal of Economic Literature*, March 1998, *36*, 11–46.

Rubinstein, Ariel, "Similarity and Decision-Making Under Risk (Is There a Utility Theory Resoltion to the Allais Paradox?)," *Journal of Economic Theory*, 1988, *46*, 145–153.

Samuelson, Larry, "Evolution and Game Theory," *Journal of Economic Perspectives*, 2002, *16*, 47–66.

Schelling, Thomas C., *The Strategy of Conflict*, Cambridge, MA: Harvard University Press, 1980.

Sen, Amartya and B. Williams, *Utilitarianism and Beyond*, Cambridge, UK: Cambridge University Press, 1982.

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

Smith, Vernon L., "Economics in the Laboratory," *Journal of Economics Perspectives*, Winter 1994, *8* (1), 113–131.

Thaler, Richard, "Toward a Positive Theory of Consumer Choice," *Journal of Economic Behavior and Organization*, 1980, *1*, 39–60.

Weibull, Jürgen, *Evolutionary Game Theory*, Cambridge MA: MIT Press, 1995.

Yellen, John E., "The Transformation of the Kalahari !Kung," *Scientific American*, April 1990, *262* (4), 96–105.

Young, H. Peyton, "An Evolutionary Model of Bargaining," *Journal of Economic Theory*, 1993, *59* (2), 145–168.