1999

Caring About Sunk Costs: A Behavioral Solution to the Hold-up Problem

Lorne Carmichael

W. Bentley MacLeod

POLITICAL
ECONOMY
RESEARCH
GROUP

PAPERS IN POLITICAL ECONOMY

Paper No. 95

"Caring About Sunk Costs: A Behavioral Solution to the Hold-up Problem"

Lorne Carmichael and
W. Bentley MacLeod

The UNIVERSITY of WESTERN ONTARIO

The Political Economy Research Group was established in the Faculty of Social Science at the University of Western Ontario in 1988. Its purpose is to foster scholarship, teaching and interdisciplinary research in political economy, with a focus on:

1. the application of economic models and methods to the study of political processes and institutions,
2. the economic impact of political processes and institutions,
3. the influence of economic factors on the formation of public policy and on institutional change,
4. the politics of economic policy making, and
5. the political, social, and economic effects of public policy.

Co-directors:
Ronald Wintrobe (Economics)
Robert Young (Political Science)

Board of Directors:
Peter Howitt (Economics)
Michael Keating (Political Science)
John N. McDougall (Political Science)
Peter Neary (History)
John Whalley (Economics)

Staff:
Yvonne Adams

For further information:
Political Economy Research Group,
Department of Economics,
Social Science Centre,
London, Ontario, Canada N6A 5C2
phone: (519) 661-2111, ext. 85231
fax: (519) 661-3666

ISSN: 1181-7151
ISBN: 0-7714-2252-0

Caring about Sunk Costs: A Behavioral Solution to the Hold-up Problem

Lorne Carmichael*
Queen's University
Kingston, Ontario K7L3N6

W Bentley MacLeod
University of Southern California
Los Angeles, CA

November 11, 1999

Abstract

Economics students need to be taught that opportunity costs are important for optimal decisionmaking but that sunk costs are not. Why should this be? Presumably these students have been making optimal decisions all their lives, and the concepts should be easy for them. We show that this behavior may be part of a cultural mechanism for dealing with the incentive problems associated with team production in a complex environment. The approach has applications to the modelling of economic behavior and to the theory of incentives in organizations.

*An eary version of this paper was presented at the Fain Conference at Brown University in October of 1998. The first author would like to thank the SSHRCC for support, while the second author is grateful for the financial support of National Science Foundation, grant SBR-9709333.

1 Introduction

The study of costs is a major part of all first year Economics courses. Students are taught that opportunity costs are important, but sunk costs are not. Sunk costs have already been paid, cannot be recovered, and should not affect current or future choices. Students often find this lesson counterintuitive, despite the fact that they have been making (presumably optimal) choices all their lives.

Even more surprising, perhaps, is that costs that have been sunk by other people also seem to affect behavior. In a well know survey Kahneman, Knetsch and Thaler (1986) asked people how much they would be willing to pay for a beer bought at a particular store and brought to them on a beach. The beer is to be delivered at a given time and at a given temperature, and is a standardized product in every way. Nonetheless, the amount customers say they will pay depends on the costs sunk by the owner of the store into the quality of his building. People are willing to pay more if the beer was bought from a fancier store.

One can find other cases where sunk costs seem to affect economic behavior. In Behavioral Finance, for example, it is a well established tenet that small investors tend to lose money because they sell their winners and hang onto their losers. Investors treat the amount they paid for a security, even though it is sunk, as a cost they would like to recover.[1]

Behavior like this does not seem to be rational in the usual economic sense. Nonetheless, we will show that there is an economic explanation for it. First, we will present an economic environment where caring about sunk costs is definitely an optimal strategy. In this model agents are presented with joint investment opportunities and must bargain over the division of the resulting surplus. People who insist that they be compensated for their sunk costs in any bargain, and who also recognize others' claims to similar compensation, have available to them a very attractive and robust mechanism for solving the hold-up problem, and the associated problems of free riding in team production environments. The mechanism is efficient in a wide class of environments and is uniquely efficient when the environment is sufficiently complex Williamson (1975).

Agents in this model are assumed, in the manner of Schelling (1980) and Frank (1988) to be able commit to walk away from a bargain if they are offered less than a certain minimum outcome. We call this amount the "fair demand" for that agent. The rules for

[1]See Chip Heath and Lang (1999) for a recent contribution, with references.

determining "what's fair" are assumed to have emerged at the cultural level and will satisfy an evolutionary stability condition. We show that there is a unique rule for determining the fair demand of each agent that is both stable and efficient, and that in this equilibrium agents' fair demands will depend on the costs they have sunk into their investments.

Note that we are using the term "fair" in a particular way. The fairness rules we derive are *not* based on normative considerations. Rather, they mark the minimum amounts that people will accept without a fight, and what they will grant to others. In our context, the statement: "That's not fair!" is a territorial claim.

We will also explore an application to the theory of incentives in organizations. The fairness rule we identify extends easily to teams of arbitrary size. We identify a role for ownership in this context, and show that organizations that pay a "fair wage" can elicit effort from workers without the use of strong incentive contracts. This may help us understand why organizations work even when explicit economic incentives seem to be weak.

The beer example above is encompassed by our model as are the results of bargaining experiments where sunk costs have been shown to affect outcomes (Borges and Knetsch (1997)). However, people also seem to care about sunk costs in situations where there is no obvious bargaining going on. In some of these situations, caring about sunk costs appears to be a bad strategy. Our explanation for this is less formal, and depends on the notions of bounded rationality and *framing* (Kahneman and Tversky (1979)). Drawing on some recent work in Cognitive Psychology, we suggest that over time the gains from being able to commit in real bargaining situations make up for the occasional losses we incur in other situations, given the biological constraints that govern our cognitive and commitment abilities. The hypothesis has testable implications, which we begin to explore.

Throughout the paper we model the commitment level of an agent as a *strategy*. This is in contrast to a behavioral approach, where concerns about sunk costs (say) might be modelled as a part of preferences. Our approach has the advantage that we can introduce behavioral realism into an economic model without sacrificing the ability to make meaningful welfare comparisons across institutional regimes. In addition, it suggests practical ways to use behavioral data to reveal "true" preferences even when observed behavior is subject to anomalies such as a concern for sunk costs. We explore these issues in a final section.

We are not the first to suggest that cultural rules for fair dealing would play an important allocative role is society, or that evolution is the appropriate framework within which to study the determination of culture. The closest antecedent to our approach in the economics literature is probably Margolis (1982), who suggested in the public goods context that efficiency can be achieved if agents bargain over the distribution of resources in proportion to what they believe has been their contribution to the public good.[2] His motivation is to understand why individuals make public contributions that are larger than can be explained with the simple self-interested model. In our case we are interested in explaining why individuals take into account sunk costs when making decisions.

2 The Model

We consider a society where risk neutral agents meet randomly in pairs. Each pairing involves a productive opportunity followed by a bargain over the division of the surplus created. In the *ex ante* stage, agents independently make a decision as to the character of an investment. In the *ex post* stage the parties meet and bargain over the resultant surplus. We assume the matching process is efficient, and will concentrate on the events that occur within a match. In particular, we will focus on the norms that might govern bargaining in the *ex post* period.

This basic framework has been studied many times before, in many variations. Alchian and Demsetz (1972) argue that any prespecified division of the surplus will lead to free riding by team members, who will all underinvest. Grout (1984) shows that the same result holds when rational agents bargain *ex post* over the division of the surplus. Holmström (1982) argues that problems arise because the sum of the marginal products of each team member adds up to more than the total product of the team, so that an entrepreneur who can break the budget balancing condition can get first best effort from team members. However, Eswaran and Kotwal (1984) argue that this sets up perverse incentives for the entrepreneur. MacLeod (1987) and Carmichael (1988) suggest that the entrepreneur has good incentives when teams of agents are put in competition with each other, but this may not always be possible. Williamson (1975), in a wide ranging discussion, argues that ex ante agreements may be subject to "hold-up", and suggests remedies including the exchange of hostages or their economic equivalents. Hart (1995) explores the role of property rights and financial structure in this context, and in a recent

[2]The specifics of Margolis (1982) are quite different from our paper, but he does propose a "Fair Share" rule that includes the following. "The larger the share of my resources I have spent unselfishly, the more weight I give to my selfish interests in allocating marginal resources."

paper Che and Hausch (1999) argue that fully efficient cooperative investments are impossible to achieve when there is the possibility of *ex post* renegotiation.

The framework of this paper is closest to that of Grout (1984), but the assumptions made about *ex post* bargaining are different. Here, bargaining over the *ex post* surplus is assumed to be *territorial*, in the sense of Carmichael and MacLeod (1998). This means that each party enters the bargain with a prespecified notion of what would be a "fair" outcome for himself. If it begins to appear that the bargain will produce less that this amount, then he can walk away and force an outcome where both sides get nothing.[3] If the offered division is acceptable to both sides, then each gets to keep what he has agreed to.

Bargaining is called "territorial" because of the parallel between the assumed bargaining behavior and an animal's defense of its territory. Agents are able to make and communicate a commitment to "fight" – i.e., engage in actions that are costly to both parties – if they are not getting what they "deserve" out of a bargain. Just as important, they are willing to agree peacefully to a division if they are getting at least this much. At the heart of the approach is the determination of the territorial boundaries – i.e. "what's fair". We model this as the outcome of an evolutionary process at the societal level. People unreflectively learn what their society considers fair, but only certain notions of fairness will survive.[4]

A bargaining strategy in our model is a mapping from the *ex post* state of the world to a "fair demand" for the agent. We will call this mapping the agent's "culture", and we will be assuming that agents share the same culture. In more familiar terms, we will be looking for symmetric equilibria in this bargaining game. Nonetheless the term "culture" serves to emphasize several important aspects of our model. First, agents have beliefs about the "appropriate" outcome in any situation. An agent who deviates from the fair demand that his culture considers appropriate may be subject to sanctions that are costly to both parties in the bargain. These are not legal sanctions, but follow from the withdrawal of cooperation by fellow members of the same society. In a world with fully rational agents the threat of these sanctions would not be credible. Further, unlike the case where there is a legal contract, the punishments can be conditional on variables that are known to the bargaining parties but are not verifiable to any third party.

[3]This is just a simple way of modelling the costs to a dispute. In general the cost to disagreement must be discrete, but need not be as painful as this.

[4]Ellingsen (1997) and Carmichael and MacLeod (1998) are other papers that use this basic bargaining framework.

More formally now, the stages in the history of a match are as follows.

1. Before each match begins all agents are anonymous. Each then learns her type $t \in T = \{1...n\}$ where the probability of being type t is given by $P(t)$, $P(t) > 0$, $n < \infty$. Let t^i be the type of agent i.

2. An agent's type determines his cost function for the investment. Each agent makes an investment $I \in \Re_+$ which requires him to pay a cost $C \geq 0$, $C \in \{C_0...C_{m-1}\}$ where $Prob\{C = C_r\} = f(C_r|I,t)$. We assume $f(C_r|I,t) > 0 \; \forall \; I \geq 0, \; t \in T$.

3. The individuals i and j are matched. Their match produces a surplus $S \in \{S_0...S_{k-1}\}$ where $Prob(S = S_r) = g\left(S_r|I^i, I^j\right) = g\left(S_r|I^j, I^i\right) > 0 \; \forall \; I^i, I^j \geq 0$.[5]

4. The surplus is observed by each party, as are the costs incurred by each. Let $\omega = \{S, C^i, C^j\}$ denote the *state* of the relationship, and let Ω denote the set of possible states.

5. Each agent decides what is a fair outcome for him from the match. Denote by $d^i = d^i(\omega) \in [0, S]$ the minimal demand of agent i. If offered anything less than this, agent i will walk away from the match.

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

$$V^i\left(d^i, d^j, \omega\right) = \begin{cases} -C^i \text{ if } S - \left(d^i(\omega) + d^j(\omega)\right) < 0 \\ d^i(\omega) + \left(S - \left(d^i(\omega) + d^j(\omega)\right)\right)/2 - C^i, \text{ if not.} \end{cases} \tag{1}$$

Given the investment I^i , a bargaining strategy for i is a function:

$$d^i : \Omega \rightarrow \Re_+. \tag{2}$$

The agent's commitment level $d^i(\omega)$ is his belief as to what is a fair outcome for him in the *ex post* bargain, and will depend on information known to him at that time. If feasible, the bargaining outcome is assumed to give each party at least an amount she considers fair – otherwise one or both would refuse to deal. If there is anything left over it is divided evenly. The rule for division of any remaining surplus is not critical – all the results follow from the assumption that an increase in the territorial claim $d^i(\omega)$, while it might increase

[5] We make the generic assumption throughout that the values of S_r and $C_{r'}$ are distinct, that is $S_r \neq C_{r'}$, for all r and r'.

the chances of conflict, will also increase the amount one gets from those matches that continue to reach agreement. This is true in a wide variety of analytic bargaining environments, including the standard Nash case where all unclaimed amounts are lost.[6]

Note that there are two aspects to a strategy – the investment level and the commitment level. The commitment level is determined *ex post* but the function $d^i(\cdot)$ is a learned, culturally dependent rule for determining what is a fair outcome from a bargain. All agents in our society are assumed to have learned the same rule for "what's fair", so while these demands may depend on the type of an agent they will not depend on his identity.

At the time the agent chooses her investment level she can anticipate the effect her investment will have on her payoff from the bargain. Optimal investments for each agent will then depend on type according to the function $\mathbf{I} : T \rightarrow \Re_+$ An overall strategy for the game is a pair of functions $\sigma = \{\mathbf{I}, \mathbf{d}\}$ where $\mathbf{I} = \{I_t\}$, $t \in T$ and $\mathbf{d} = \{d_t(\omega)\}$, $t \in T$, $\omega \in \Omega$.

If $\omega = (S, C^i, C^j)$ then we define $\overline{\omega} = (S, C^j, C^i)$. Given the strategy of agent $j, \sigma^j = \{\mathbf{I}^j, \mathbf{d}^j\}$, for agent i of type t^i, whose has made investment I_{t^i}, the probability of state ω is defined by:

$$H\left(\omega|I^i_{t^i}, t^i, \mathbf{I}^j\right) = \sum_{t^j \in T} g\left(S|I^i_{t^i}, I^j_{t^j}\right) f(C^i|I^i_{t^i}, t^i) f(C^j|I^j_{t^j}, t^j) P(t^j). \tag{3}$$

We can now define the expected payoffs to player i against player j in the game as

$$U(\sigma^i, \sigma^j) = \sum_{t^i \in T} \sum_{t^j \in T} \sum_{\omega \in \Omega} V^i\left(d^i_{t^i}(\omega), d^j_{t^j}(\overline{\omega}), \omega\right) H\left(\omega|I^i_t, t, \mathbf{I}^j\right) P(t^i) P(t^j). \tag{4}$$

Given that the game is symmetric, the payoff to player j is simply $U(\sigma^j, \sigma^i)$.

We assume that there is a unique vector of strictly positive investments maximizing social welfare in any match. Individuals, however, will choose investment levels to maximize their individual expected payoffs given their cultural beliefs. Our focus is on the character of this cultural rule for fair division $d_t(\omega)$. What cultural rules will be stable? Is there a rule that can induce individuals to make efficient investments?

We know very little about the actual dynamics of cultural change, and we will not attempt to model them explicitly. Rather, we follow Maynard Smith (1982) and use the static concept of evolutionary stability. Evolutionarily stable strategies are known to be

[6] In general the territorial claim must be real, but the rule for the division of what is left over can be quite general.

rest points for a number of different specific dynamics (see Weibull (1995) or Malaith (1998) for a review of this literature). The basic idea is that successful cultural rules will be those that lead to good outcomes for individuals in environments where the majority of agents follow the rule. More formally, if a small population playing a different cultural strategy "invades" an incumbent population, an agent from the invading population must do worse in the new mixed population than the incumbent agents. This is equivalent to:

Definition 1 *A strategy $\tilde{\sigma}$ is an Evolutionarily Stable Strategy (ESS) iff:*

1. (Nash Equilibrium) $U(\tilde{\sigma},\tilde{\sigma}) \geq U(\tilde{\sigma}',\tilde{\sigma})$ for all $\tilde{\sigma}' \in \Sigma$
2. (Stability) If $U(\tilde{\sigma},\tilde{\sigma}) = U(\tilde{\sigma}',\tilde{\sigma})$, $\tilde{\sigma} \neq \tilde{\sigma}'$, then $U(\tilde{\sigma},\tilde{\sigma}') > U(\tilde{\sigma}',\tilde{\sigma}')$

The first requirement is that the incumbent strategies form a Nash equilibrium. The second is the stability condition. Suppose a small "mutant" population enters, and its members do as well against the incumbents as the incumbents do against themselves. Evolutionary Stability requires that all members of the invading population be strictly worse off in the new mixed population than agents from the original population. It follows that they must do worse against each other than they do against members of the incumbent population. It is also clear that any strict Nash equilibrium must be evolutionarily stable.

While in general there may be many Evolutionarily Stable Strategies, we will be looking for strategies that are both stable and efficient. One reason to focus on efficient equilibria is simply that it makes our result that sunk costs matter more provocative. Caring about sunk costs is not rational behavior, so it is interesting that such behavior can nonetheless be part of an efficient and stable equilibrium strategy. However, there are also several reasons why one might expect more efficient rules to prevail in an evolutionary process, assuming of course that stable and efficient rules exist in the first place. For one, Sobel (1993) shows that cheap talk will allow agents to find more efficient equilibria.[7] Other authors, including Binmore (1994) have argued that humans evolved in small competing groups. each with a distinct shared culture. In this case the notion of group selection might be appropriate for the study of the cultural rules that govern the informal division of resources within a defined social group. Group selection leads directly to greater efficiency at the shared cultural level.[8]

[7] The equilibrium concept must be weakened to neutral stability since the strategy now includes the (perhaps meaningless) messages sent.

[8] We require that the cultural rule be a standard ESS, so that group selection enters only as a refinement.

We shall show that the following "fair share" rule is, under appropriate conditions, the unique efficient evolutionarily stable equilibrium in this model.

Definition 2 *The* fair share *rule is defined by:*

$$d^i(\omega) = \textit{sunk costs paid by } i + \textit{an equal share of the net surplus,} \tag{5}$$
$$= C^i + \frac{S-(C^i+C^j)}{2}, \omega = \{S, C^i, C^j\}. \tag{6}$$

This rule has the following characteristics:

1. It is evolutionarily stable.

2. Agents believe it is fair that they be compensated for costs they have sunk, and, equally important, they also believe it is fair that other people be compensated for the costs that these others have sunk.

3. There are no disagreements in equilibrium, so the rule is *ex post* efficient.

4. The rule provides first best incentives for *ex ante* investment, even when optimal investments depend on worker type.

5. Since the rule is independent of an agent's type, it can be implemented even when information about worker types is unavailable *ex post*.

The rest of this section will establish these claims and present conditions under which the *fair share* rule is unique. Notice first that there are no disagreements *ex post* when for every $\omega \in \Omega$, $d^i(\omega) + d^j(\bar{\omega}) = S$, a condition that the fair share rule satisfies. Notice this implies $d^i(\omega) = S + d^j(\bar{\omega})$, and since the right hand side is independent of agent $i's$ type this implies that *every* rule that is *ex post* efficient must be type independent.

Let us introduce some notation that will prove useful below. Since the state space is finite, then each state can be indexed: $\Omega = \{\omega_\ell\}$, where $\omega_\ell = \{S_r, C_s^i, C_v^j\}$, with $\ell = m^2r + ms + v + 1$, which runs from one to km^2. Hence we can write the demands and parameters as vectors:

$$\hat{d} = \begin{bmatrix} d_1 \\ d_2 \\ \cdot \\ d_{km^2} \end{bmatrix}, \text{ and } \hat{S} = \begin{bmatrix} S_1 \\ S_2 \\ \cdot \\ S_{km^2} \end{bmatrix}, \tag{7}$$

where $d_\ell = d(\omega_\ell)$, S_ℓ is given by $\omega_\ell = \left\{S_\ell, C_\ell^i, C_\ell^j\right\}$ and $\hat{C}^i, \hat{C}^j$ are defined in a similar fashion. Let

$$\hat{P} = \begin{bmatrix} P(1) \\ P(2) \\ \cdot \\ P(n) \end{bmatrix} \tag{8}$$

and let $\mathbf{H}(\mathbf{I}^i, \mathbf{I}^j)$ be the $n \times km^2$ matrix with $t\ell$ entry given by $\mathbf{H}_{t\ell}(\mathbf{I}^i, \mathbf{I}^j) = H(\omega_\ell | I_t^i, t, \mathbf{I}^j)$, where t indexes the agent type and ℓ indexes the commonly observed state.

Social welfare is defined by:

$$W\left(\mathbf{I}^i, \mathbf{I}^j\right) = \hat{P}\mathbf{H}\left(\mathbf{I}^i, \mathbf{I}^j\right)\left(\hat{S} - \hat{C}^i - \hat{C}^j\right). \tag{9}$$

Let $\mathbf{J}_{t\ell}(\mathbf{I}^i, \mathbf{I}^j) = \partial H(\omega_\ell | I_t^i, t, \mathbf{I}^j) / \partial I_t^i$. We have assumed the existence of a unique optimum, say $(\mathbf{I}^{*i}, \mathbf{I}^{*j})$. Since the model is symmetric, then $(\mathbf{I}^{*j}, \mathbf{I}^{*i})$, gives the same payoff, from which we conclude $\mathbf{I}^{*i} = \mathbf{I}^{*j}$, and hence we may denote the social optimum by $\mathbf{I}^*$, which must satisfy the first order conditions:

$$\mathbf{J}(\mathbf{I}^*, \mathbf{I}^*)\left(\hat{S} - \hat{C}^i - \hat{C}^j\right) = \hat{0}. \tag{10}$$

Since agent $i's$ investment affects only her costs, it follows that $\mathbf{J}(\mathbf{I}^*, \mathbf{I}^*)\hat{C}^j = \hat{0}$, from which we conclude that the first order conditions characterizing the first best are also given by:

$$\mathbf{J}(\mathbf{I}^*, \mathbf{I}^*)\left(\hat{S} - \hat{C}^i\right) = \hat{0}. \tag{11}$$

It is now easy to see that:

Proposition 3 *The* fair share *rule is evolutionarily stable and ensures efficient* ex ante *investments.*

Proof. We have assumed there exists a unique interior optimum which is characterized by the first order conditions (10). Under the *fair share* rule the demand of player i is given by $\hat{d}^i = \left(\hat{S} - \hat{C}^i - \hat{C}^j\right)/2 + \hat{C}^i$, and hence given that the investment strategy for player j is the optimal one, $\mathbf{I}^*$, the first order condition for player i is:

$$\mathbf{J}\left(\mathbf{I}^i, \mathbf{I}^*\right)\left(\hat{d}^i - \hat{C}^i\right) = \mathbf{J}\left(\mathbf{I}^i, \mathbf{I}^*\right)\left(\hat{S} - \hat{C}^i - \hat{C}^j\right)/2 \tag{12}$$

$$= \hat{0}. \tag{13}$$

Since the optimal investment strategy is unique, player $i's$ best response is to choose $\mathbf{I}^i = \mathbf{I}^*$, and hence the optimal investment strategy forms a Nash equilibrium. The assumption of uniqueness also implies that this is a strict Nash equilibrium and hence we may conclude that the *fair share* rule results in an efficient ESS. ■

The efficiency properties of the *fair share* rule are not hard to understand. Each of the agents to the bargain knows that he will receive a predermined share of the overall net surplus in every *ex post* state. Everyone shares the costs of investment in the same proportion that they share the revenues, and, at the margin, behaves just like a residual claimant. (Note that $S_{ij} - (C_i + C_j)$ need not be positive – if there is not enough surplus to cover investment costs, agents believe it is fair that the net losses be shared evenly.)

There are, of course, many possible rules for the *ex post* efficient division of a surplus. However, we can also show that so long as investments can affect the size of the surplus, any cultural division rule that leads to efficient ex ante investments must depend on something more than just the surplus itself. To see this, suppose that demands are *not* sensitive to sunk costs. This implies that $d^i(\omega) = d^i(\bar{\omega})$ for all $\omega \in \Omega$. When combined with the efficiency requirement, $d^i(\omega) + d^j(\bar{\omega}) = S$, this implies that $d^i(\omega) = S/2$. However, as Grout (1984) as shown, such a rule results in suboptimal investment, as summarized in the following proposition.

Proposition 4 *Suppose that the division rule is both* ex post *efficient and independent of sunk costs. Then at any ESS demands are given by $d_t(\omega) = S/2$. If investment affects the distribution of returns, then assuming the existence of an interior equilibrium, investment is inefficient.*

Proof. We have already shown that if sunk costs are disregarded, then the only *ex post* efficient rule is the equal division rule. Now consider investment incentives.

The payoff to agent i under the equal division rule is:

$$U\left(\sigma^i, \sigma^j\right) = \hat{P}^T \mathbf{H}\left(\mathbf{I}^i, \mathbf{I}^j\right)\left(\hat{S}/2 - \hat{C}^i\right). \tag{14}$$

Any ESS must be a Nash equilibrium, implying that for each type, I_t is chosen to maximize $i's$ payoff. Moreover, given the symmetry of the model $\mathbf{I}^i = \mathbf{I}^j$ in equilibrium. The ESS investments, $\mathbf{I}^{ESS}$must satisfy:

$$\mathbf{J}\left(\mathbf{I}^{ESS}, \mathbf{I}^{ESS}\right)\left(\hat{S}/2 - \hat{C}^i\right) = \hat{0}. \tag{15}$$

We suppose that investment increases the surplus over the relevant region, and hence $\mathbf{J}(\mathbf{I}^*, \mathbf{I}^*)\hat{S} > \hat{0}$, from which we conclude by comparing (15) and (10) that under the equal division rule investment incentives are inadequate. ■

The final issue concerns the uniqueness of the *fair share* rule. It is not in general unique. For example, if investments do not affect the size of the surplus, then any division of the surplus that is independent of costs, such as the equal split rule, results in efficient investment. It is the sharing of the *returns* from investment that generates inefficiency in the case of the equal split rule, and hence a necessary condition for uniqness is $\mathbf{J}(\mathbf{I}^*, \mathbf{I}^*)\hat{S} \neq \hat{0}$. This condition is implied by the following rank condition that we will show is also sufficient to ensure that the *fair share* rule is unique.

Definition 5 *The payoffs satisfy the* full rank condition *if at the efficient investment level* $\operatorname{rank}\mathbf{J}(\mathbf{I}^*, \mathbf{I}^*) = km^2 - m - 1$.

In the proof of the following proposition we show that the rank of $\mathbf{J}(\mathbf{I}^*, \mathbf{I}^*)$ can be no greater than $km^2 - m - 1$. To ensure that it is at least this large, it is necessary that there are at least $km^2 - m - 1$ different types of agents in T. The simplest non-trivial example entails $k = m = 2$, and hence at least 5 types. The important economic insight here, clearly, is that the uniqueness of the *fair share* rule depends upon sufficient unobserved heterogeneity in agent characteristics. Note that $\left\{\hat{S}, \hat{C}^i, \hat{C}^j\right\}$ are generically independent since this holds whenever the possible values of S and C are distinct.

Proposition 6 *Suppose that unique efficient investment level satisfies the full rank condition. The* fair share *rule is then the unique division rule resulting in an efficient ESS.*

Proof. Given that the investment of agent i has no effect on the cost for agent j then investment I_t^i has no effect on the probability of the cost C_r being realized by agent j. Let $E_r = \{\omega \in \Omega | C^j = C_r\}$. This implies that for each $t \in T$, and for each value of costs $C_r \in \{C_0, ..., C_{m-1}\}$:

$$\sum_{\omega_\ell \in E_r} \mathbf{J}_{t\ell}\left(\mathbf{I}^i, \mathbf{I}^j\right) = 0. \tag{16}$$

Given that there are m such events, this implies that $rank\,\mathbf{J}(\mathbf{I}^i, \mathbf{I}^j) \leq km^2 - m$. Let N^j denote the space spanned by the vectors e^r, where $e_\ell^r = 1$ if $\omega_\ell \in E_r$ and zero otherwise. Notice that $\hat{C}^j = C_0 e^o + C_1 e^1 + ... + C_{m-1} e^{m-1}$, and hence $\hat{C}^j \in N^j$. Let N^i denote the corresponding space from the perspective of agent j, for which $\hat{C}^i \in N^i$.

Since S_r and $C_{r'}$ take on distinct values, then $\left\{\hat{S}, \hat{C}^i, \hat{C}^j\right\}$ are linearly independent and $\hat{S} - \hat{C}^i \notin N^j$. Thus given that at the efficient investment level, $\mathbf{J}(\mathbf{I}^*, \mathbf{I}^*)\left(\hat{S} - \hat{C}^i\right) = \hat{0}$, and $\hat{C}^j \in N^j$, the rank condition implies that the nullspace of $\mathbf{J}(\mathbf{I}^*, \mathbf{I}^*)$ is spanned by

$\left\{\left(\hat{S}-\hat{C}^i-\hat{C}^j\right), N^j\right\}$. Now suppose that $\hat{d}^i$ is a division rule resulting in an efficient ESS, then it must solve the first order condition:

$$\mathbf{J}\left(\mathbf{I}^*, \mathbf{I}^*\right)\left(\hat{d}^i-\hat{C}^i\right)=0, \tag{17}$$

which from the rank condition implies that $\hat{d}^i-\hat{C}^i=\alpha\left(\hat{S}-\hat{C}^i-\hat{C}^j\right)+v^j$, where $v^j \in N^j$. Similarly, $\hat{d}^j-\hat{C}^j=\alpha\left(\hat{S}-\hat{C}^i-\hat{C}^j\right)+v^i$. Efficiency implies that $\hat{d}^i+\hat{d}^j=\hat{S}$, from which we conclude:

$$(1-2\alpha)\hat{S}=(1-2\alpha)\left(\hat{C}^i+\hat{C}^j\right)+v^j-v^i. \tag{18}$$

The right hand side is in the span of $\{N^i, N^j\}$, while the left hand side is not and hence $\alpha=1/2$. Thus $v^j-v^i=\hat{0}$, but given that $N^i \cap N^j=\hat{0}$, we conclude that $v^i=v^j=\hat{0}$ and we are done. ∎

The required notion of fairness we derive is exactly what was described in the introduction. People care about sunk costs, in that they bargain as if they want to be compensated for expenses they have already paid. As well, and equally important, they think it is fair to accept less from a bargain in order to compensate others for the investments that they have made.

3 Informational Issues

The framework that we have used in this paper involves a finite number of agent types, surplus levels, and individual costs, and relies also on a full support assumption that keeps each agent's type unknown *ex post.* This has allowed us to establish that the efficient "fair demand" mechanism is unique in a particular context. However, the mechanism works in such a transparent fashion, essentially by making all parties net residual claimants, that it should be efficient in a much wider set of circumstances.

For example, consider the case examined recently by Che and Hausch (1999). In an illustrative example, they consider a situation where a buyer has a valuation for a good given by $v(e)$, $v(0)=0$, $v'(0)>1$, where e is the cost of the input of the seller. The seller's valuation is zero. This is a purely "cooperative" investment, in that the seller gains nothing directly from his actions. Nonetheless the first best outcome is clearly for the seller to provide the input level e^*, where $e^*=\arg\max\{v(e)-e\}$ and for the good to be transferred to the buyer. The levels of v and e are observable to the parties ex post, but are not contractible. The authors distinguish between two cases – one where renegotiation of a contract is not allowed, and one where it cannot be prevented if it is in the interests of the two parties.

When renegotiation is not allowed there is a simple contract that will achieve first best. The parties agree to give the buyer an option to buy at the price $p = v(e^*)$. Since he observes $v(e)$ *ex post*, the buyer will only purchase the good if $v(e) \geq p = v(e^*)$, and the seller will therefore have the incentive to set $e = e^*$. When renegotiation cannot be prevented, however, the buyer can reject the option but offer to buy at a lower price. Che and Hausch (1999) show that the subgame perfect outcome will be one where the contract is irrelevant. Rational bargaining will set a transfer price of $v(e)/2$ and the seller will choose e to maximize $v(e)/2 - e$. The first best cannot be achieved. The authors go on to show in a more general model that the first best is unattainable when investments are cooperative and renegotiation is allowed.

The "fair demands" mechanism we characterize above will also achieve first best in this context. The seller will invest knowing that his choice of e will be observed ex post by the buyer. He enters the *ex post* bargain with the credible demand that he be paid at least $p = e + (v(e) - e)/2$. The buyer, having made no investments himself, agrees that a fair price will indemnify the seller for his sunk costs e and buys at this price.[9] The seller chooses e to maximize $(v(e) - e)/2$, and the efficient level is chosen.

In a sense the "fair demand" mechanism acts like a commitment not to renegotiate, because agents are able to walk away if their fair demands are not met. Why can't the parties just agree *ex ante* that the only fair price is $p = v(e^*)$, and trade at this price or not at all? It might seem that all we have done is exhibit another contract that will work if renegotiation is not allowed, and there was never any claim in Che and Hausch (1999) that the option mechanism was unique.

The important difference between the sunk cost mechanism and the option scheme is in their requirements for information. Both mechanisms assume that *ex post* information about actions and outcomes are available to each party, although it is not verifiable. However, the option mechanism requires as well that *ex ante* information about the seller's

[9] Note that the buyer and seller are identified as such *ex ante*, and so other efficient outcomes are possiple. In general any price $p = e + \alpha(v(e) - e)$, $\alpha \in (0, 1)$ can be an efficient "fair demand". The parameter α is a "fair return" for the seller and must be established *ex ante*, perhaps just by the supply and demand of sellers. We do not pursue this here, since our point is that regardless of α agents care about and respect sunk costs.

investment technology be available to the buyer, so that the correct option price can be determined and incorporated into an explicit contract.

Consider what is probably the simplest possible extension to incomplete information. Suppose that the buyer's *ex post* valuation is given by $v(e) \pm \theta$, $\theta > 0$. Let the probability of the good state be π. Then the optimal effort level of the seller is unchanged, and we assume that $v(e^*) - \theta - e^* > 0$. To ensure trade in each state the option price must be $p = v(e^*) - \theta$. If the seller provides e^* then she gets the expected return $p - e^* > 0$. However, she can also provide effort e_{low} and sell only in the good state. The effort level e_{low} solves $p = v(e_{low}) + \theta$ and gives the return $\pi(p - e_{low})$. Given that $e^* > e_{low}$, for sufficiently large π the seller will not provide optimal effort.

The sunk costs mechanism works in this case, and in cases where even less is known about the technology of production. Suppose the seller's technology was drawn from the set $\psi = \{v_t(e)\}$. The buyer does not know t beforehand, and need not even know anything about the set of possibilities ψ. All he needs to know is the outcome, the seller's costs, and that a fair return for the seller will respect those costs – i.e., $p^{fair} = e + (v - e)/2$. As well, the sunk costs mechanism is decentralized – it requires no third parties at all, even to enforce a simple fixed price option contract.

If information about costs and outcomes is verifiable, then an explicit mechanism that compensates rational people for their costs could also achieve first best. If costs cannot be observed at all, then clearly the *fair share* mechanism cannot work. However, in those cases where *ex post* cost information is available to the people involved, and so long as cultural norms can be enforced by credible threats of punishment, fair bargainers have a very robust mechanism for achieving efficient joint investments. It works in a wide variety of circumstances and requires very little advance knowledge. The mechanism is ideal for agents who can make commitments but who live in a world they do not completely understand.

4 Ownership

Models of organizations normally assume that incentive problems are mitigated through the use of incomplete contracts. These contracts are designed to induce rational people to behave in a more efficient manner. As noted by Macauley (1963), however, the use of explicit incentive schemes is quite rare. Our approach suggests that workers and managers, by virtue of their common culture, will have available behavioral strategies that worked in the family and in the playground, and might be usefully adapted to the workplace. This section works out an example.

In proving our main result we showed first that demands cannot be type dependent, which requires that the net surplus be divided equally. However, the share of the net surplus given to each agent does not affect her incentives, so long as it remains positive. Thus a concern for sunk costs will provide an exact solution to multilateral investment, or "team production" problems. For a team of size N, the fair demand $d_i(\omega) = C_i + [S - (\sum C_j)]/N$ will lead to optimal investment by each agent. The possibility of disagreement is always present, but the team's budget is balanced in equilibrium. *Ex ante*, members need to know only the expected effect of their own investments on the bargaining surplus. *Ex post*, the parties must be aware of the costs that each has paid and of the size of the surplus, but this information need not be verifiable – i.e. there is no need for outside parties such as the court to enforce contracts. No single person at any time needs to know the overall team investment technology.

While the solution is not affected in principle by the size of the team, in practice bargaining among a great many team members will be impractical. However, such an organization may still rely on the "cultural capital" of its members to help solve internal incentive problems. Many of the insights of Alchian and Demsetz (1972) apply to this case.

Suppose that the surplus is produced by an asset K, that requires an "entrepreneur" to make the investment I^E, at an ex post cost of C^E. In addition, the asset requires N workers to make investments denoted by I^i. The total surplus produced by the relationship is S and is assumed to be observable *ex post*. At the time negotiation occurs the investment cost of a worker is known to that worker and the entrepreneur, as is the size of the surplus (i.e. the revenue of the team). However C^E is unobserved by workers, and hence the bargaining solution above cannot be implemented. The state of the match to a worker is $\omega^i = \{C^i, S\}$, while to an entrepreneur it is $\omega^E = \left\{C^E, C^1 ... C^N, S\right\}$. The entrepreneur bargains with each worker individually.

In this case we can approach the first best arbitrarily closely with strategies that make the person with the entrepreneurial input the major residual claimant. Consider a scheme where each worker gets an ε share of profits. For small ε each worker claims:

$$d^i(\omega) = \text{sunk costs} + \text{small share of the net surplus} \tag{19}$$

$$= C^i + \varepsilon \left(S - \sum_{j=1..N} C^j \right) \tag{20}$$

while the entrepreneur claims:

$$\begin{aligned} d^E(\omega) &= \text{most of the net surplus} && (21) \\ &= (1 - N\varepsilon)\left(S - \sum_{j=1..N} C^j\right) && (22) \end{aligned}$$

Using the arguments from Proposition 6, it is straightforward to derive the following:

Proposition 7 *For every $\varepsilon > 0$ the worker invests efficiently, while as $\varepsilon \to 0$ the entrepreneur's investment becomes arbitrarily close to the efficient level. The demands $d(\omega) = \left\{d^i(\omega), d^E(\omega)\right\}$ form an ESS that approaches the efficient allocation as $\varepsilon \to 0$.*

This result is very similar in spirit to the one proposed by Alchian and Demsetz (1972). These authors argued that having a team member who is central to all contracts cuts down dramatically on the costs of bargaining. The entrepreneur was also needed to monitor inputs and reward or punish workers accordingly, and had to be assigned the residual in order to make sure he did his job properly. It was never very clear, however, how he was to motivate workers since once he had the residual he was constrained to a fixed overall wage bill. One possibility is that he might pay an efficiency wage, and fire those who fail to invest. To set this wage properly he would need to know the technology of worker investment, however. Carmichael (1983) suggested that he might put agents in competition with each other, but this also requires information about technologies, and might lead workers to avoid helping activities or to engage in outright sabotage.

What we have shown above is that by paying a "fair wage" – i.e., a wage that respects the worker's sunk costs cost of investment – the entrepreneur need only add a very small share of firm profits to ensure efficient behavior. In practice the worker's costs need to be observable, and this brings up issues that are not unlike those investigated by Holmström and Milgrom (1991) in the context of output based incentive schemes. It is important to emphasize, however, that in our context costs do not need to be verifiable to influence the outcome of bargains between a worker and her employer.

Some important costs may be easy to measure such as the opportunity cost of the worker's time, which can be covered by a fixed hourly wage. The opportunity cost of effort, which has been a focus of most previous work in incentive theory, may not be so important to measure accurately. Given that a worker has chosen his occupation and is on the premises, the marginal opportunity cost of effort may be zero – what else is she going to do? Under these conditions the observed use of a fair hourly wage and a minimal profit sharing plan can achieve first best, even though the free rider problem would seem insurmountable. On the entrepreneur's side, even though his costs are not observed, investment is close to efficient so long as the level of profit sharing is small.

5 Cognitive Issues

Up to this point agents' concern for sunk costs has been optimal behavior in its given context. But people seem to be influenced by sunk costs even in situations where such behavior does not seem optimal. People hang on to their losing investments too long, refuse to fold a weak poker hand when they have a lot of their "own" money in the pot, and stay to watch the end of movies they don't like in order to "get their money's worth". Assuming first that there is an explanation for this behavior, and that it deserves the attention of economists, can our model account for it?

In animal studies, where the assumption of bounded rationality is unavoidable, it is well understood that nonoptimal behavior will arise when an organism is moved out of the environment where the behavior originated. Many newly hatched birds for example, will "imprint" on the largest nearby being and follow it wherever it goes. In normal times this behavioral strategy will lead the young bird towards food and away from danger. Out of context, when a bird born in captivity follows its human keeper, the behavior seems quaint and humorous. However, given the biological constraints on the bird's possible cognitive abilities, this is clearly a strategy that has withstood many years of evolutionary pressure and served these birds very well.

In this section we explore the idea that a similar phenomenon might account for people's concern for sunk costs, even in circumstances where such a concern is inappropriate. To do this requires a major departure from standard models of economic decisionmaking. In particular, we must accept as a hypothesis the idea that mechanisms for instinctive decisionmaking and rational decisionmaking exist side by side in the human brain. We do not attempt to model this formally in this paper. Rather, we will argue that the idea is plausible, we will outline some of its testable implications, and we will point out what we think is a major methodological advantage over alternatives that account for anomalous behavior by adding variables to the utlitity function.

5.1 Reason or Instinct?

The idea that there might be independent systems in the human brain that govern our decisionmaking receives support from recent work in Cognitive Psychology.[10] As LeDoux (1996) discusses in some detail, humans have at least two physiologically distinct systems that take environmental cues or stimuli and transform them into actions. What he calls "the high road" works through the cerebral cortex, involves reasoning, and takes time. The "low road" allows individuals to respond very quickly to changes in their environment. It 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. Also, once a response has been learned it is difficult to unlearn.

For example, if someone is driving and sees 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. People who move to a country where people drive on the other side of the road will eventually learn to react safely, but there is a period where reactions cannot be trusted.

Given that overall behavior is being controlled by different and autonomous systems, one of which is designed to react before the situation has been completely understood, it makes sense that the systems might sometimes be in conflict. 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 knew he could not be hurt, and he had consciously commanded his head to stay close to the glass. Most of us would do the same. The key point is that there are times when we behave in certain predictable ways without immediate understanding and in spite of our conscious efforts to the contrary.

This research makes it at least plausible that something similar could be the source of framing effects like a concern for sunk costs. At a conscious level, a person sitting at a football game in the rain understands that the price he paid for the ticket is gone, and he should really be inside getting warm. But he remains in order to "get his money's worth" from the ticket. This stubborn conviction would be useful were there really a bargain going

[10] This is also the basis of much recent work in Evolutionary Psychology. See Cosmides and Tooby (1994) and the references therein.

on over the services he is receiving and the price he should pay. It is not useful while he is sitting in the rain, but we hypothesize that given its value in real bargaining situations, and given the limitations on our ability to recognize situations that require commitment, the strategy is beneficial overall.

5.2 Implications

If our hypothesis is true, and people care about sunk costs because they have "framed" their situation as a bargain even when it is not, then we should be able to predict how they will behave under alternate frames. This is the source of testable implications.

For example, many students in an undergraduate class will say that they would stay to watch the end of a $10.00 movie thay didn't like, rather than leave to meet a friend at a bar next door. Others will say they would leave, and argue that the money is gone anyway so they have nothing further to lose. If the hypothetical price of admission is raised to $100.00, however, more students will say they would stay. People care about sunk costs. Most interesting, however, is that if the hypothetical situation is changed so that the $100.00 admission was a donation to charity, virtually all students say they would leave the movie.

This simple demonstration, easily replicable in any first year class, suggests quite strongly that we care about sunk costs in situations where we are concerned about getting fair value for our money. Once the concern for fair treatment is gone, we behave in a more rational manner.

As a second example, let us return to the case where people were asked how much they would be willing to pay for a beer brought to them on a beach. Under our model it is clear that even though subjects were asked to reveal their "willingness to pay" they were not revealing anything about their preferences. They were revealing was a strategy. Based on information about the costs the owner has sunk into his premises, respondents were trying to predict the minimum price he would accept, and were choosing their offered price accordingly. They wanted a beer, but they did not want to pay too much for it.[11]

To find out how much someone "really wants" a beer, our approach suggests that the experimenter should make sure that her subjects are not framing the situation as an investment/ bargain.[12] For example, the experimenter could place a cold beer on one table

[11] It seems clear, if this interpretation is correct, that costs need not be observed with perfect accuracy for our model to be useful. By the same token, a storeowner should not expect to be able to raise his prices for making investments (e.g. high quality plumbing) that his customers do not observe.

[12] We suspect that similar problems would arise even if the beer were auctioned off on the beach. Bidders

and a ten dollar bill on another, and see which one the subject decides to pick up. On a hot day we suspect many subjects would pick up the cold beer, even though they would never dream of paying this much for one. There is no reason to suspect the decision to depend on the store from which the beer was puchased.

5.3 Welfare

There is now a large body of work in "behavioral" decision theory. This work progresses by specifying a form for the utility function so that when maximized subject to the usual constraints, realistic behavior is generated. The best known 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.

The approach generates the behavior it was designed to explain in a straightforward way. However, it is not so useful as a basis for making welfare judgements. This problem has been used to justify the rationality assumption in economic models even when it is clear that the behavioral implications of rationality are false (Myerson (1999)).

The "bicameral" model of the brain we propose puts instinctive and rational decision making on an equal footing. Both the "high road" and the "low road" map information about the environment into actions. Both systems occasionally make mistakes. And, since both systems implement strategies based on an underlying fitness ordering that is assumed to be well behaved,[13] we can use the model as a basis for making welfare judgements.

We have already seen one aspect of this. In the previous sections we were able to compare the sunk costs mechanism with rational bargaining and the option mechanism on the basis of their investment incentives and their various demands on information. One could surely model people's concern for sunk costs by placing them as a "reference point" (Kahneman and Tversky (1979)) into the utility function. But then preferences themselves would change when costs are sunk. This makes suspect the whole concept of economic efficiency, since to evaluate an investment strategy we must compare welfare across states where people have different preferences.

are competing with each other, but they are also aware of how much "excess profit" the auctioneer is making.

[13]The distinction between utility and fitness is also examined in Bergstrom (1996).

The bicameral model also gives us some guidance when people are behaving in what appears to be a suboptimal way. For example, there is a role for investment counsellors – they can help their clients by convincing them to sell their losers. Under the standard behavioral approach these investors are just maximizing their preferences. They "really care" about their losers, and will be worse off if they sell them. There is no way for an investment counsellor to tell them they are mistaken.

6 Conclusions

In a complex world people know a great deal more about what has happened in the past than they do about what may happen in the future. This makes attractive any mechanisms that allow for *ex post* settling up of accounts. In the absence of enforceable contracts that can hold people responsible for their actions in the past, previous work has argued that sequential rationality will prevent settling up, and that inefficiencies are inevitable.

In a world where bargaining agents can commit to territories that are contingent on information that is available *ex post*, there is another instrument –a cultural mechanism – that can be applied to the problem of attaining efficient investments in joint relationships. When beliefs about fair treatment in a bargain include compensation for the sunk costs of investments that are germane to the match, efficient levels of investment can be sustained. These "fair demands" are evolutionarily stable, and if the evolutionary process favors efficient institutions there is some likelihood that they will arise and be robust. This may help to explain how organizations are able to prosper even when explicit economic incentives for team members seem too weak.

In a more general sense, this research is clearly concerned with exploring and extending a particular explanation for some of the departures from rationality that seem to appear in and outside the laboratory. Underlying the story is a "bicameral" model of human decision making that suggests anomalous behaviors arise because they are functional in situations that are similar to the one being examined. The approach has several advantages.

1. It provides a satisfying explanation for the behavior based on more primitive truths about the structure of our brains and the fact that we and our culture are the products of evolution.

2. The theory is testable. It allows us to make predictions about behavior in experiments and in situations that have not yet been investigated.

3. It allows the economic researcher to use the standard tools of welfare economics to define efficiency, evaluate existing institutions and make policy suggestions. This is not possible when preferences themselves change in response to previously sunk investments.

References

Alchian, Armen and Harold Demsetz, "Production, Information Costs, and Economic Organization,," *American Economic Review,*, December 1972, *62* (5), 777–795.

Bergstrom, Theodore C., "Economics in a Family Way," *Journal of Economic Literature*, December 1996, *34*, pp. 1903–1934.

Binmore, Ken, *Playing Fair: Game Theory and the Social Contract*, Vol. 1, Cambridge, MA, USA: MIT Press, 1994.

Borges, Bernhard F. J. and Jack L. Knetsch, "Valuation of Gains and Losses: Fairness and Negotiation Outcomes," *International Journal of Social Economics*, 1997, *24*, 265–281.

Carmichael, H. Lorne, "The Agents-agents problem: Payment by relative output," *Journal of Labor Economics*, 1983, *1*, 50–65.

Carmichael, H. Lorne., "Incentives in Academics: Why Is There Tenure?," *Journal of Political Economy*, 1988, *96*, 453–72.

Carmichael, H. Lorne and W. Bentley MacLeod, "Fair Territory: Bargaining, Property Rights, and Framing Effects," Technical Report, Queen's University and University of Southern California 1998.

Che, Yeon-Koo and Donald B. Hausch, "Cooperative Investments and the Value of Contracting," *American Economic Review*, March 1999, *89* (1), 125–47.

Cosmides, Leda and John Tooby, "Better than Rational: Evolutionary Psychology and the Invisible Hand," *American Economic Review, Papers and Proceedings*, May 1994, pp. 327–32.

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

Eswaran, Mukesh and Ashok Kotwal, "The Moral Hazard of Budget-Breaking," *RAND Journal of Economics*, Winter 1984, *15* (4), 578–581.

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.

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

Grout, Paul, "Investment and Wages in the Absence of Binding Contracts: A Nash Bargaining Approach," *Econometrica*, March 1984, *52* (2), 449–460.

Hart, Oliver D., *Firms, Contracts and Financial Structure*, Oxford, UK: Oxford University Press, 1995.

Heath, Steven Huddart Chip and Mark Lang, "Psychological Factors and Stock Option Exercise," *Quarterly Journal of Economics*, May 1999, *64* (2), 601–28.

Holmström, Bengt, "Moral Hazard in Teams," *Bell Journal of Economics*, 1982, *13*, 324–40.

—— **and Paul Milgrom**, "Multi-Task Principle-Agent Analysis: Incentive Contracts, Asset Ownership, and Job Design," *Journal of Law, Economics, and Organization*, 1991, *7*, 24–52.

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

—— **, Jack L. Knetsch, and Richard Thaler**, "Fairness as a Constraint on Profit Seeking: Entitlements in the Market," *American Economic Review*, September 1986, *76* (4), 728–741.

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

Macauley, Stewart, "Non-contractual Relations in Business: A Preliminary Study," *American Sociological Review*, 1963, *55*, 55–69.

MacLeod, W. Bentley, "Behavior and the Organization of the Firm," *Journal of Comparative Economics*, 1987, *11*, 207–220.

Malaith, George, "Do People Play Nash Equilibria? Lessons from Evolutionary Game Theory," *Journal of Economic Literature*, 1998, pp. 1347–1374.

Margolis, Howard, *Selfishness, Altruism, and Rationality:A Theory of Social Choice*, Cambridge: Cambridge University Press, 1982.

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.

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

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

Sobel, Joel, "Evolutionary Stability and Efficiency," *Economics Letters*, 1993, *42* (2-3), 301–312.

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

Williamson, Oliver E., *Markets and Hierarchies: Analysis and Antitruct Implications*, New York: The Free Press, 1975.