Caring About Sunk Costs: A Behavioral Solution to Holdup Problems with Small Stakes

Lorne Carmichael
Queen's University

W. Bentley MacLeod
University of Southern California

Economics students need to be taught that opportunity costs are important for optimal decision making 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 caring about sunk costs can help agents achieve efficient investments in a simple team production environment. Furthermore, the solution we propose is uniquely efficient if the environment is sufficiently complex. Hence, in addition to explaining contract form and ownership (Williamson, 1975; Hart, 1995), studies of the holdup problem may also provide insights into observed behavior in day-to-day bilateral bargaining problems.

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, even though (we assume) they have been making 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-known survey Kahneman, Knetsch, and Thaler (1986) asked people whether they thought it would be fair for a store to increase its prices under various circumstances. People were unwilling to accept a higher price due to higher demand, but they would accept higher prices due to higher input costs. For example, a hardware store that increased its prices for snow shovels after a serious snowstorm was behaving badly, but it was perfectly acceptable for this store to raise prices

An early version of this article was presented at the Fain Conference at Brown University in October 1998. The authors thank the referees and Ken Binmore for their thoughtful comments. The first author would like to thank the SSHRCC for support, while the second author is grateful for the financial support of the National Science Foundation (grant no. SBR-9709333).

if the cost of its stock had gone up, even though this stock has already been bought and is sitting on the shelves.

There are other examples. Bewley (1997) found that employees are more willing to accept changes in compensation arising from changes in the firm's fortunes than in response to changes in labor market demand. In financial markets, it is often claimed that small investors lose money because they sell their winners and hang on to their losers (Chip Heath and Lang, 1999). Investors treat the price paid for their investment, even though it is sunk, as a cost they would like to recover.

Economists for the most part have shown little theoretical interest in this behavior, arguing that any concerns for an "appropriate" price will give way eventually to the forces of supply and demand. We would make this argument as well, so long as the market is fluid. However, when prices are set in a situation of bilateral monopoly, we will show that there is a simple *economic* explanation for the way that the outcome of a bargain might respond to the costs sunk by the various parties. Caring about sunk costs can lead to economically efficient outcomes when agents are faced with joint investment opportunities and must then bargain over the division of the resultant surplus. Moreover, when there is sufficient variation in underlying characteristics, we will show that this convention is the *unique* efficient division rule.

Our model builds on Williamson's (1975) observation that in a complex world, when there are relationship-specific investments, it is often very difficult to prespecify ex ante the terms of an ex post relationship. What the parties may do instead is lay down some rules that will affect their future interactions. The recent theory of the firm, starting with Klein, Crawford, and Alchian (1978), Williamson (1985), and Grossman and Hart (1986), explores how parties can, with incomplete explicit contracts or a careful allocation of property rights, provide incentives for investment. However, first best is often not attainable.

Our primary concern in this article is with situations where individuals are unable to write explicit contracts. One case, consistent with the retailing examples above, would be where agents must make their investments before they meet their partner for the match. The standard approach in this literature has been to assume a reduced form bargaining game for the ex post stage—usually the Nash bargaining solution. Here, as Grout (1984) has shown, individuals underinvest in relationship-specific investments since in the end they will receive less than their marginal contribution to the surplus. The result is similar to the familiar "team production problem" of Alchian and Demsetz (1972).

In this article we modify the bargaining stage of the Grout (1984) model to incorporate the insight of Schelling (1980) and Frank (1988) that individuals may be able to make emotional demands that push the outcome in their favor. In a similar vein, Crawford (1982) builds a formal model of bargaining impasses when individuals have different costs of backing down and this information is private to each person. Since our goal is to understand the character of bargaining outcomes rather than the existence of breakdowns, we restrict our attention to the case where information at the bargaining stage is symmetric.

In this game, any division of the surplus is a Nash equilibrium, and newly matched individuals face a coordination problem in deciding how to play. We suggest that social norms of fair division may have evolved to solve this coordination problem in a useful way. Each agent will enter a bargain with a preconceived notion of what would be a "fair" outcome for him. Most important, each also has the ability to walk away from the match and force an outcome where each side gets nothing rather than accept something he considers unfair.

Note that "fairness" in this context has nothing to do with normative considerations. A bargainer's demand for "what's fair" marks the minimum amount that she will accept without a fight and the amount that she will grant to others. The demand is like a territorial claim. A norm of fair division in a society is like a stable assignment of territory to each agent.[1]

We are not the first to suggest that norms for fair division play an important allocative role in society. For example, Hayek (1982) argues that observed social norms arise as efficient solutions to social dilemmas. Margolis (1982) has used this idea to suggest that individuals develop "fair share" rules for the provision of funds to public projects. Ellickson (1991) has used this approach to explain the trespass rules for cattle in Shasta County, California. He finds that in the absence of clear legal rules (or in some cases in spite of the existence of a legal rule), individuals develop social norms of conduct that efficiently resolve potentially difficult bargaining problems. Recently Binmore (1994, 1998) has developed an evolutionary theory of utilitarian notions of fairness.[2]

Binmore's theory, like many of the previous theories, focuses on the role that fairness plays in reducing conflict rather than on why a particular allocation is more efficient. Our contribution is simply to point out that another social convention—caring in a particular way about the costs you and others have sunk into a relationship—may also be an efficient solution to a social dilemma, that is, the holdup problem. We also show that when there is sufficient diversity in individual characteristics, this social convention is the

[1] Akerlof (1980) considers a somewhat different model of social customs. In his framework, those who deviate from the custom are punished by others. The punishers carry through out of fear they will be punished for failing to punish, and so on. We could establish our results in this framework as well—the key is simply that it pay individuals to conform to the customs of their society. However, the norms we study are enforced by the sense of outrage that greets transgressors, and thus can be implemented even when there are no third parties to observe a transaction.

[2] Ken Binmore observes that our theory has some links to equity theory in psychology (see section 4.4 of Binmore [1998]). Equity theory allows for the possibility that individuals receive in proportion to their worthiness; however, as Binmore observes, the notion of worthiness is never well defined. Our contribution can be viewed as providing an economic basis for worthiness based on sunk investments, though the rule we find is different from the one that equity theorists suggest.

only rule for dividing the surplus that will always lead to efficient ex ante investments. This means that while there may be other solutions that work in specific cases (see, e.g., Hart, 1988; Che and Hausch, 1999.), this one will work in those cases and many more besides. Thus it is ideal for situations where the parties do not know enough about each other to design a specific contractual mechanism that would take account of their personal characteristics or information. We will provide some examples after the model is developed.

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.

A bargaining strategy in our model is a mapping from the ex post state of the world to a "fair demand" for the agent. The ex post state is known to the parties, but need not be verifiable. "What's fair" can therefore depend on things that would be impossible to establish in a court of law. Nonetheless, an agent who deviates from the norm may be subject to costly sanctions.

More formally, the stages in the history of a match are outlined below. The model has a discrete character that will be useful when we consider the role of complexity.

1. Before each match begins all agents are anonymous. Each then learns her type $t \in T = \{1 \dots n\}$, 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 $C \geq 0$, $C \in \{C_0 \dots C_{m-1}\}$, where $\Pr\{C = C_r\} = f(C_r|I, t)$. 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 $S \in \{S_0 \dots S_{k-1}\}$, where $\Pr(S = S_r) = g(S_r|I^i, I^j) = g(S_r|I^j, I^i) > 0$ $\forall I^i, I^j \geq 0$.[3]
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.

[3] 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'.

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

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

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

$$d^i\colon \Omega \to \Re_+. \tag{2}$$

The agent's commitment level $d^i(\omega)$ is his belief in 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 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. We could also just pick randomly an outcome from the set that both parties consider fair.[4]

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 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 the type according to the function $\mathbf{I}\colon T \to \Re_+$. 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 $\bar{\omega} = (S, C^j, C^i)$. Given the strategy of agent j, $\sigma^j = \{\mathbf{I}^j, \mathbf{d}^j\}$, for agent i of type t^i, who made investment I_{t^i}, the probability of state ω is defined by

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

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

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

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

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

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 the cultural norm for fair division $d_t(\omega)$. We will propose a rule shortly that can induce individuals to make efficient investments. We will show that this rule is stable and will explore some of its properties. Then we will argue that it seems to capture some aspects of actual behavior.

The dynamics of cultural change are complex, and we will not attempt to model them explicitly. Rather, we will simply show that it is in the interest of individuals in the population to conform to the rule for fair division, so long as everyone else does, and that the preference is strict. It is well known that strict Nash equilibria are also evolutionarily stable (Maynard Smith, 1982; van Damme, 1991) and are therefore rest points for a number of different specific dynamics (see Weibull, 1995, or Malaith, 1998, for a review of this literature).

We are now in a position to state the main result of our article. The following "fair share" rule is, under appropriate conditions, the unique efficient equilibrium in this model.

Definition 1. The "fair share" rule is defined by

$$d^i(\omega) = \text{sunk costs paid by } i + \text{an equal share of the net surplus,} \tag{5}$$

$$= C^i + \frac{(S - C^i - C^j)}{2}, \tag{6}$$

where $\omega = \{S, C^i, C^j\}$. $$\tag{7}$$

This rule has the following characteristics:

1. The strategies form a strict Nash equilibrium.
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 $\ell = m^2 r + ms + v + 1$, which runs from one to km^2. Hence we can write the demands and parameters as vectors:

$$\hat{d} = \begin{bmatrix} d_1 \\ d_2 \\ \cdot \\ d_{km^2} \end{bmatrix}, \text{ and } \hat{S} = \begin{bmatrix} S_1 \\ S_2 \\ \cdot \\ S_{km^2} \end{bmatrix}, \tag{8}$$

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

$$\widehat{P} = \begin{bmatrix} P(1) \\ P(2) \\ \cdot \\ P(n) \end{bmatrix} \tag{9}$$

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

$$W(\mathbf{I}^i, \mathbf{I}^j) = \widehat{P}^T \mathbf{H}(\mathbf{I}^i, \mathbf{I}^j)(\widehat{S} - \widehat{C^i} - \widehat{C^j}). \tag{10}$$

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:

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

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

$$\mathbf{J}(\mathbf{I}^*, \mathbf{I}^*)(\widehat{S} - \widehat{C^i}) = \widehat{0}. \tag{12}$$

It is now easy to see that

Proposition 2. The *fair share* rule is a strict Nash equilibrium and ensures efficient ex ante investments.

Proof. We have assumed there exists a unique interior optimum that is characterized by the first-order conditions of Equation (11). Under the fair share rule the demand of player i is given by $\widehat{d^i} = (\widehat{S} - \widehat{C^i} - \widehat{C^j})/2 + C^i$.

Given that the investment strategy for player j is the optimal one, $\mathbf{I}^*$, the first-order condition for player i is

$$\mathbf{J}(\mathbf{I}^i,\mathbf{I}^*)(\hat{d}^i-\widehat{C}^i)=\mathbf{J}(\mathbf{I}^i,\mathbf{I}^*)(\widehat{S}-\widehat{C}^i-\widehat{C}^j)/2 \tag{13}$$

$$=\hat{0}. \tag{14}$$

Since the optimal investment strategy is unique, player i's best response is to choose $\mathbf{I}^i=\mathbf{I}^*$, and the optimal investment strategy forms a Nash equilibrium. The assumption of uniqueness also implies that this is a strict Nash equilibrium. At this point it is clear that any other demand will get player i strictly less and hence we may conclude that the fair share demands form an efficient strict Nash equilibrium. ■

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 predetermined 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.) It is also clear that this rule will produce efficient investments in a more standard framework where, say, the surplus is a continuous increasing function of the costs of investment by each party.

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 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) has shown, such a rule results in suboptimal investment.

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 that is independent of costs, such as the equal split rule, results in efficient investment. There are also contractual solutions when the investment is made by one party only (Hart and Moore, 1988) or when the efficient investment levels are known ex ante and there is third-party enforcement. But there are many other cases where the parties do not know enough about each other ex ante to design one of these arrangements, or where the benefits from the match are too small to make a contractual solution worthwhile. The strength of the fair share rule is that it works in simple environments and it also works in situations that are much more complex or where the parties have much less information. Our next proposition shows that if the environment is sufficiently complex, in that there is enough unobserved heterogeneity among agents, then the fair share rule is the only rule that generates efficient investments.

It is clear that a necessary condition for uniqueness is $\mathbf{J}(\mathbf{I}^*, \mathbf{I}^*)\widehat{S} \neq \hat{0}$, that is, investments must affect the size of the surplus or else the equal split rule will work. 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 3. 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 nontrivial example entails $k = m = 2$, and hence at least five types. The important economic insight here is that the uniqueness of the fair share rule depends on sufficient unobserved heterogeneity in agent characteristics. Note that $\{\widehat{S}, \widehat{C^i}, \widehat{C^j}\}$ are generically independent since this holds whenever the possible values of S and C are distinct.

Proposition 4. 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 strict Nash equilibrium.

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 $C_r \in \{C_0, \ldots, C_{m-1}\}$,

$$\sum_{\omega_\ell \in E_r} \mathbf{J}_{t\ell}(\mathbf{I}^i, \mathbf{I}^j) = 0. \tag{15}$$

Given that there are m such events, this implies that $\operatorname{rank} \mathbf{J}(\mathbf{I}^i, \mathbf{I}^j) \leq km^2 - m$. Let N^j denote the space spanned by the vectors e^r, where $e^r_\ell = 1$ if $\omega_\ell \in E_r$ and zero otherwise. Notice that $\widehat{C^j} = C_0 e^0 + C_1 e^1 + \cdots + C_{m-1} e^{m-1}$, and hence $\widehat{C^j} \in N^j$. Let N^i denote the corresponding space from the perspective of agent j, for which $\widehat{C^i} \in N^i$.

Since S_r and $C_{r'}$ take on distinct values, then $\{\widehat{S}, \widehat{C^i}, \widehat{C^j}\}$ are linearly independent and $\widehat{S} - \widehat{C^i} \notin N^j$. Thus given that at the efficient investment level, $\mathbf{J}(\mathbf{I}^*, \mathbf{I}^*)(\widehat{S} - \widehat{C^i}) = \hat{0}$, and $\widehat{C^j} \in N^j$, the rank condition implies that the null space of $\mathbf{J}(\mathbf{I}^*, \mathbf{I}^*)$ is spanned by $\{(\widehat{S} - \widehat{C^i} - \widehat{C^j}), N^j\}$. Now suppose that $\hat{d}^i$ is a division rule resulting in an efficient Nash equilibrium, then it must solve the first-order condition

$$\mathbf{J}(\mathbf{I}^*, \mathbf{I}^*)(\hat{d}^i - \widehat{C^i}) = 0, \tag{16}$$

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

$$(1 - 2\alpha)\widehat{S} = (1 - 2\alpha)(\widehat{C^i} + \widehat{C^j}) + v^j - v^i. \tag{17}$$

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. ■

In order for the fair share rule to work it is clear that the costs of investment must be observable to the two parties. This does limit its applicability. However, so long as costs are known ex post, this last result shows that the rule is remarkably robust. In particular, since the rule works for a wide variety of agent types, it means that agents have very little information ex ante about the people they are dealing with. Each party can invest with confidence knowing that they will share the net surplus, whatever it happens to be. They do not need to monitor each other's investments, and have no information about what the other's optimal investment should be or what the resultant surplus should be. The rule is ideal for a world where people know what has happened in the past but do not have enough information to accurately predict what may (or should) happen in the future.

3. Some Examples

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 they have made. This simple ex post bargaining rule can help agents achieve first-best investments even when they have little ex ante information about the people they are dealing with.

The closest thing to a direct test of this model, so far as we are aware, was conducted by Borges and Knetsch (1997). In a series of experiments, subjects were asked to arbitrate a division of a \$10.00 gain between two people. In one treatment neither party had incurred any costs, and the overwhelming choice was the 50/50 split. In the next treatment one of the parties had incurred a cost of \$2.00 in order to participate. This time 76% of the arbitrators thought a 60/40 split would be fair, so that each party could go home with \$4.00. The fact that the \$2.00 cost was sunk did not seem to matter. It is interesting that in a third treatment where one of the parties had given up a \$2.00 opportunity in order to participate, as opposed to paying a cost to participate, 83% of the arbitrators ignored this and considered the 50/50 split to be appropriate. Very similar results were obtained when the arbitrators were asked to divide up a \$10.00 loss between the parties.

Therefore, even though lost opportunities and sunk costs are equivalent economic situations, the individuals in these bargaining experiments respond quite differently, and in particular take into account sunk costs, while ignoring lost opportunities. These experiments did not test bargaining behavior per se, but they do speak to common notions of what is a fair outcome in a bargain when the costs to each party are known ex post. Sunk costs matter in precise accordance with the fair share rule.

This rule also seems to be in action in the retail pricing sector. In a well-known 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.

Another example occurs in the way that fresh fish is priced in restaurants. Menu prices at a restaurant do not vary with daily changes in demand. Yet even though there are transaction costs associated with changing restaurant prices, the amount for a fresh fish dinner is often set each day based on the (sunk) cost of the fish bought that morning. This means that the restaurant owner can invest in the best quality of fish knowing that his customers will pay the extra cost of a more expensive species.

Retailing examples like these fit the notion of complexity that was essential in proving the uniqueness of the fair share rule. Despite the monopolistic aspects of these examples, there is no way that a store owner could set up an ex ante agreement with each of his potential customers to pay more if he sinks money into the quality of his premises or his product. He has no idea who his customers will be at the time he makes his investment, and they will have no idea how much it costs to do the appropriate thing. However, in a culture that recognizes the fair share rule, a retailer can invest knowing that his customers will pay a higher price if they perceive his costs are higher.[5]

4. Conclusion

Studies of the holdup problem (Williamson, 1975; Klein et al., 1978; Grossman and Hart, 1986) have been very important for our understanding of ownership and contract form. In this article we have shown that the study of the holdup problem can also provide some insights into the fair division problem in bilateral bargaining situations. Moreover, the "fair share" solution that we have derived has the property that an individual's demand is increasing in their sunk investment and decreasing with their trading partner's investment, a result that is consistent with personal experience, some experimental evidence, and with the survey evidence of Kahneman, Knetsch, and Thaler (1986).

We have derived the form of the fair share rule in a simple context to highlight the structure of the efficient norm. There are clearly some issues that deserve further examination. For example, our model does not have any conflict in equilibrium, which might arise if we extended the model to include asymmetric information, as in Crawford (1982). Second, we do not

[5] Of course there are limits, since the fair share rule applies to situations of bilateral monopoly and there is always some competition. As well, since costs are not directly observed by customers, the retailer may decide to invest in things that are expensive in appearance only.

present an explicitly dynamic analysis of the evolution of the division rule, in part because there is still no widespread agreement regarding the form that such a dynamic should take.[6] As well, it is clear from our retailing examples that sunk costs can influence the outcome of a bargain even when these costs are not precisely observed, which is something we did not model explicitly. Nonetheless, we have shown that a simple, realistic, and fully efficient solution to the holdup problem exists in a fairly general context. We hope this demonstration will provoke further extensions.

References

Akerlof, G. A. 1980. “A Theory of Social Custom of Which Unemployment may be One Consequence,” 94 *Quarterly Journal of Economics* 749–75.

Alchian, A., and H. Demsetz. 1972. “Production, Information Costs, and Economic Organization,” 62 *American Economic Review* 777–95.

Bewley, T. F. 1997. “A Depressed Labor Market, as Explained by Participants,” unpublished manuscript, Yale University.

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

———. 1998. *Game Theory and the Social Contract.* Volume 2: *Just playing.* Series on Economic Learning and Social Evolution. Cambridge, MA: MIT Press.

Borges, B. F. J., and J. L. Knetsch. 1997. “Valuation of Gains and Losses: Fairness and Negotiation Outcomes,” 24 *International Journal of Social Economics* 265–81.

Che, Y.-K., and D. B. Hausch. 1999. “Cooperative Investments and the Value of Contracting,” 89 *American Economic Review* 125–47.

Chip Heath, S. H., and M. Lang. 1999. “Psychological Factors and Stock Option Exercise,” 64 *Quarterly Journal of Economics* 601–28.

Crawford, V. P. 1982. “A Theory of Disagreement in Bargaining,” 50 *Econometrica* 607–37.

Ellickson, R. C. 1991. *Order Without Law: How Neighbors Settle Disputes.* Cambridge, MA: Harvard University Press.

Frank, R. H. 1988. *Passions Within Reason.* New York: W. W. Norton.

Grossman, S. J., and O. D. Hart. 1986. “The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration,” 94 *Journal of Political Economy* 691–719.

Grout, P. 1984. “Investment and Wages in the Absence of Binding Contracts: A Nash Bargaining Approach,” 52 *Econometrica* 449–60.

Hart, O. D. 1988. “Incomplete Contracts and the Theory of the Firm,” 4 *Journal of Law, Economics, & Organization* 119–39.

———. 1995. *Firms, Contracts and Financial Structure.* Oxford: Oxford University Press.

———, and J. Moore. 1988. “Incomplete Contracts and Renegotiation.” 56 *Econometrica* 755–85.

Hayek, F. A. 1982. *Law, Legislation and Liberty.* London: Routledge and Keagan Paul.

Kahneman, D., J. L. Knetsch, and R. Thaler. 1986. “Fairness as a Constraint on Profit Seeking: Entitlements in the Market,” 76 *American Economic Review* 728–41.

Klein, B., R. Crawford, and A. Alchian. 1978. “Vertical Integration, Appropriable Rents, and the Competitive Contracting Process,” 21 *Journal of Law and Economics* 297–326.

Malaith, G. 1998. “Do People Play Nash Equilibria? Lessons from Evolutionary Game Theory,” 36 *Journal of Economic Literature* 1347–1374.

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

[6] See Binmore (1994, 1998) and Young (1998) for an impressive start toward a dynamic theory of social conventions and norms.

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

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

van Damme, E. 1991. *Stability and Perfection of Nash Equilibria*, 2nd ed. Berlin: Springer-Verlag.

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

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

———. 1985. *The Economic Institutions of Capitalism.* New York: Free Press.

Young, P. H. 1998. *Individual Strategy and Social Structure: An Evolutionary Theory.* Princeton, NJ: Princeton University Press.