Can Contract Theory Explain Social Preferences?

W. Bentley MacLeod
Columbia University
New York, NY

December 21, 2006

During the last several decades, a growing body of laboratory research has shown that human subjects do not always choose to maximize material payoffs. Economists following the lead of psychologists Daniel Kahneman and Amos Tversky (1979) and Mathew Rabin (1993) have built on such research to suppose that individuals are concerned about the distribution of material rewards between themselves and others as well as with personal payoffs. (Ernst Fehr and Klaus Schmidt (1999)) An alternative approach, suggested by Elizabeth Hoffman, Kevin McCabe and Vernon Smith (1996) argues that subjects in the laboratory perceive the situation as one with the potential for future interaction, and their behavior is motivated by a preference for reciprocity.

Both approaches capitalize on the power of psychology to enhance our understanding of economic exchanges between particular people in specific laboratory conditions. Yet the converging interest of psychology and economics in systems of rewards and social preferences has depth and breadth, dating at least as far back as Adam Smith in his *The Theory of Moral Sentiments* (1759). The convergence today is marked by precise research on the degree of the effectiveness of incentives, but also in quantifying the role of trust, fairness, honesty, and other attributes in helping people in such important goals as curtailing harmful drinking, smoking, and eating habits and learning to interact more rewardingly with partners, friends, and colleagues.

Economists have supposed that individuals maximize rewards not because they believe that people do so in every case, but because the maximizing supposition provides a useful unified model of behavior that may guide the design of better economic institutions. The model helps not only in rationalizing the profit motive, but in explaining how even successful institutions may be destabilized by ignoring the common human desire to maximize possessions–if that desire leads to malfeasance, theft, or general corruption.

In this note, however, I suggest that economics may also learn more from psychology about positive behaviors–such as the complex dynamics of honesty, fairness, and trust. Rather than insist that people are endowed with a stable set of unchanging preferences, we may ask instead how small modifications in preferences can lead to significant improvements in economic performance.

Specifically, I argue that contract theory can be a starting point for a larger enquiry into exactly how preferences might be modified to enhance efficiency. I consider two substantive themes. First I show how contract theory can provide a foundation for a theory of fairness. Second, I show that even a small taste for honesty leads to large increases in cooperation and performance.

Fairness is an especially crucial theme because notions about it are often invoked in determining whether or not an individual has breached an agreement. For example, an authoritative and widely used treatise on contract law by Edward A. Farnsworth has twenty index entries under the heading "Good Faith and Fair Dealing". The Uniform Commercial Code of the United States requires all parties to a contract to act in "goodfaith,"defined to mean "the observance of reasonable commercial standards of fair dealing. "

The problem is that "fair" is never precisely defined. In the next section, I outline a model of fairness developed with Lorne Carmichael (Carmichael and MacLeod (2003)). Following the insights of Williamson (1975) and Hart (1995), we suggest that one can develop a concept of fairness based upon the simple idea that it is optimal to reward sunk investment, and hence "fair" bargains should take this into account. This may help explain the observations of Kahneman, Knetsch and Thaler (1986) and many others who have shown that individuals exhibit an *endowment effect.* The benefit of contract theory is that it provides a basis upon which to make *predictions* regarding the nature of the preferences that lead to the endowment effect.

My second key theme is reciprocity in exchange. Beginning with the work of Ernst Fehr, Simon Gachter and George Kirchsteiger (1997) the experimental literature demonstrates that in the context of exchange with incomplete contracts, if one person does a good deed, that may lead to a positive reciprocal response. An implication of this behavior is that in finitely repeated situations parties achieve a level of cooperation that is higher than the cooperation that would be achieved if all players simply maximized their material payoffs.

We teach our children that they should abide by their obligations. This implies that they should experience disutility from breaching an agreement. ("You're grounded.") In section II, I consider a modification of an idea about agreements introduced by Hart and Holmström (1987). They observe that cooperation can be sustained if one supposes that some individuals are inherently honest but most are not. A drawback of the Hart and Holmstrom model, however, is that complete honesty and dishonesty are rather extreme cases–as noted above, most individuals place at least some value on material rewards weighed against perfect honesty. Hence it is safe to say that complete honesty is relatively rare in practice. ("I need the car to get to my study group.") Therefore, as Ken Arrow observes: "Trust is an important lubricant of a social system."[1]

The question I address in section II is whether a small amount of trust is sufficient to support cooperative behavior. I show that this is indeed the case. Therefore it follows that if a person can be taught to experience a small amount of remorse when breaking an agreement, it is possible to "socialize" the person to function well in institutions that achieve a high degree of cooperation. Moreover, the theory provides some guidance in predicting the form that a contract should take in such cases, and in suggesting *how* preferences should be modified.

I. Fairness

Students in first year economics 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 at least some optimal choices all their lives.

Indeed we know, surprisingly perhaps, that even costs that have been sunk by other people seem to affect behavior. In a well-known survey, Kahneman et al. (1986) asked people whether they thought it would be fair for a store to increase its prices under various circumstances. They found that individuals are unwilling to accept a higher price due to higher demand, but they would accept higher prices due to higher input costs. For example, for a hardware store to raise prices for snow shovels after a serious snowstorm is considered gouging, but it was considered perfectly acceptable for the store to raise prices if the cost of its stock had gone up, even though the stock had already been bought and was sitting on the shelves.

In Carmichael and MacLeod (2003) we introduce a model in which 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 resulting surplus. We assume the matching process is efficient, and concentrate on the events that occur within a match. In particular, I focus on the norms that might govern bargaining in the *ex post* period.

A bargaining strategy in the 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 for a pair are outlined below.

1. Two agents $i \in \{1,2\}$ make investments I_i that affect the distribution of costs realized, and the surplus from trade via a state ω. It is assumed that investment I_i affects costs her own costs, $C^i(\omega)$, but not her partner's costs $C^j(\omega)$, $j \neq i$. Both investments can affect surplus from future trade, denoted $S(\omega)$. Investment is not in and of itself costly–all costs enter explicit via C^i. Costs, but not investments, are observable by both parties.

2. Each agent plays a Nash demand game and chooses his or her fair share $d^{\ i}$ of S. Agent i's payoff is:

$$V^i(I_1, I_2, \omega) = \begin{cases} -C^i(\omega) \ \text{ if } S(\omega) - (d^i + d^j) < 0 \\ d^i + \left(S(\omega) - d^i + d^j\right)/2 - C^i(\omega), \ \text{ if not.} \end{cases} \tag{1}$$

If demands are less than the total surplus, then each agent gets their demand, an equal share of the remaining surplus, less their costs. If demands exceed total surplus, then each agent gets none of the surplus, but is responsible for their out-of-pocket costs. Carmichael and MacLeod (2003) show that 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:*

$$d^i = \textit{ sunk costs paid by } i \ + \ \textit{an equal share of the net surplus,} \tag{2}$$

$$= C^i + \frac{\left(S - C^i - C^j\right)}{2}. \tag{3}$$

At the *ex post* stage, any division of the surplus is efficient, and as Binmore (1994) highlights, different solutions can be interpreted as different possible social norms. Yet, most economic exchange entails some element of investment. When a plumber makes a home visit, he typically will not set the price in advance, even though upon arrival he may not have an alternative job available. The homeowner might try to take advantage of the plumber and offer a lower price, yet the consequence is likely to be quite predictable. The plumber is likely to be upset and simply leave. Such behavior, while *ex post* inefficient, does ensure that renegotiation will not occur, and the plumber is able to earn a competitive return upon his investment in skills.

In such a case, the theory is providing terms to the agreement that parties would have signed had they met before making a sunk investment. In addition to providing a normative benchmark, the results are consistent with the effects documented by Kahneman et al. (1986). They find that buyers may be unwilling to patronize sellers who increase prices due to shifts in market demand that generate a pure rent to the seller. But they are willing to accept price increases arising from increases in costs.

II. Reciprocity and a Taste for Trustworthiness

A contract is an instrument between two parties that specifies each party's obligations. In traditional societies a contract, or more precisely a *covenant*, often had an elaborate set of rituals associated with its formation. Parties might be expected to take an oath, or place their seal upon a document.[2] All such rituals were imperfect mechanisms to ensure that parties would understand their agreement and feel a sense of *obligation* to perform. Yet if the rituals had no effect, then it is unlikely that parties would have spent so much time and effort on them. Rather, rituals can be understood as conscious efforts to change each other's preferences to instill a sense of obligation that a covenant will be executed as agreed.

Realistically, all enforcement is imperfect, and must rely upon some level of trustworthiness. The question is how much. I shall show that if upon entering an agreement a party feels or experiences some loss of utility from breaching an agreement, this greatly enhances the performance of the relationship. In particular, in a finitely repeated relationship, parties can achieve close to the first best if one party has some preference for performing the contract.

Consider a buyer and seller who meet f times (the frequency of the relationship) over a unit length of time (say a year). For simplicity, there is no discounting, and periods are denoted by $\tau = 0, 1, 2, .., f-1$ corresponding to time $\tau\delta$ after the start of the relationship, where $\delta = 1/f$ is the length of one trading period. The goal of the relationship is to trade a good of flow quality $q_\tau \geq 0$ in exchange for a flow price p_τ in period τ. The payoff of the buyer in period t is $U_t^B = \sum_{\tau=t}^{f-1} \delta \left(v\left(q_\tau\right) - p_\tau \right)$, while the seller has payoff $U_t^S = \sum_{\tau=t}^{f-1} \delta \left(p_\tau - q_\tau \right)$, where $v\left(q\right)$ is the value of quality q. This value function is assumed to satisfy $v\left(0\right) = 0, v' > 0, v'' < 0$. The efficient level of quality is $q^* > 0$, and it is given by the unique solution to $v'\left(q^*\right) = 1$.

Let us suppose that the best alternative for the seller is zero, while the buyer has all the bargaining power. This is modelled by supposing each period the buyer makes a take-it-or-leave-it offer to the seller for one period of trade. It is assumed that there are no enforceable contracts. Rather, the role of the contract is to specify the obligations of each party in advance of each period.

A. Contracts: Pay in Advance Contract

Two contracts are possible. The first is the pay-in-advance contract denoted by $\{p, q\}$. At the beginning of the period, the buyer pays a flow price p, and the seller promises to deliver a flow of q at the end of the period.[3] Further suppose that this contract induces social preferences in the seller such that she or he will suffer a flow utility loss of u_c should they breach the agreement. It is assumed that $u_c > 0$, but is significantly smaller than the efficient level of quality q^*.

Given this, we can solve for the optimal contract by backwards induction. Let $\hat{U}_t^B$ and $\hat{U}_t^S$ be the equilibrium utility of the buyer and seller respectively in period t. The dynamic programing algorithm can be used to derive the optimal contract in period t :

$$\max_{q_t,p_t} U_t^B = \delta\left(v\left(q_t\right)-p_t\right)+\hat{U}_{t+1}^B \text{subject to:}$$
$$\text{IR: } \delta\left(p_t-q_t\right)+\hat{U}_{t+1}^S \geq 0,$$
$$\text{IC: } \delta\left(p_t-q_t\right)+\hat{U}_{t+1}^S \geq \delta\left(p_t-u_c\right).$$

The first constraint is the requirement that the seller get at least his or her alternative payoff. The second constraint incorporates the seller's preference to perform. If the seller does not perform, the buyer ceases all trade because they believe the seller untrustworthy. At the same time, the seller suffers a utility lost δu_c. The solution is as follows. The relationship stops in the last period, $f-1$, and therefore $\hat{U}_f^S = \hat{U}_f^B = 0$. From the IC constraint this implies that $q_{f-1} \leq u_c$. The IR constraint implies that the buyer will not offer more than the seller's alternative payoff, and hence we conclude $\hat{U}_{f-1}^S = 0$ and $p_{f-1} = q_{f-1} = u_c$. Working backwards, it follows that the unique solution is $p_t = q_t = u_c$ and $\hat{U}_t^S$ for all $t = 0, ..., f-1$. Thus, under this solution, quality is completely determined by the seller's preference to perform.

B. Contracts: Payment upon Delivery

If instead of having the seller reciprocate with quality, suppose that the seller delivers the good, and then sends an invoice to the buyer. This contract is denoted by $\{q_t, p_t\}$. Should the seller deliver $\hat{q}_t < q_t$, then the buyer has no obligation to make any payment. If at any time during the relationship the buyer does not reciprocate with a payment p_t, then the seller would choose to cease the relationship. In this case the dynamic programming problem becomes:

$$\max_{q_t,p_t} U_t^B = \delta\left(v\left(q_t\right)-p_t\right)+\hat{U}_{t+1}^B \text{ subject to:}$$
$$\text{IR: } \delta\left(p_t-q_t\right)+\hat{U}_{t-1}^S \geq 0,$$
$$\text{ICS: } \delta\left(p_t-q_t\right)+\hat{U}_{t-1}^S \geq -\delta\hat{q}_t+\hat{U}_{t-1}^S, \text{ if } \hat{q}_t < q_t$$
$$\text{ICB: } \delta\left(v\left(q_t\right)-p_t\right)+\hat{U}_{t-1}^B \geq \delta\left(v\left(q_t\right)-u_c\right).$$

It will still be the case that the seller gets zero utility, but now ICS implies that she or he will agree to choose any quality satisfying $q_t \leq p_t$. It is also interesting to note that in this formulation the seller does not feel any obligation to perform if they are not paid. This formally captures the Uri Gneezy and Aldo Rustichini (2002) observation that explicit prices (or fines) result in individuals conforming behavior to the letter of the agreement.

Solving this backwards, the solution is characterized by a price and quantity that rise as we move backwards from the endpoint until we reach an amount close to the social optimum. Figure 1 illustrates the efficiency of the relationship as a function of the period. Notice that the simulated output is close efficiency early in the relationship, then efficiency decreases toward the end.

On this graph I also plot the results reported in Armin Falk, David Huffman and W. Bentley MacLeod (2007), from an experimental labor market in which the buyer can offer bonus pay to the seller. In that case, the efficiency is relatively flat, then dips down near the end. Hence, the time structure is very similar to the one predicted by the simple model with limited trust. Though, as we can see, one does not obtain full efficiency.

III. Conclusions

From the behavioral economics literature we know that there is large variation in social preferences across individuals, both within the same society and across societies. Henrich, Boyd, Bowles, Camerer, Fehr, Gintis and McElreach (2001) and Jean Ensminger (2004) document some of this variation. They also show that an increase in market integration results in an increase in the importance of other factors regarding preferences. Thus, no stable set of preferences that can be easily estimated across a wide variety of situations exists.

As Robert Gibbons (1997) emphasizes, the evidence that individuals respond to material incentives and rewards is much more consistent. To generalize further, any mechanism (including recognizing the benefits of fairness and honesty) that link's a person's actions to something they value may be an incentive. As parents we spend a good deal of time teaching our children to value always telling the truth, treating others with respect and fairness, and cultivating other skills that foster cooperation. Similarly, firms often make heavy investments in time and effort to teach employees the value of respect and honesty in dealing with customers.

This note argues that contract theory can help guide us regarding *how* such preferences may be effectively modified. In situations where parties make investments that enhance the productivity of a match, an efficient notion of fairness is one that fully compensates individuals for sunk investments.

In the context of a relational contract, I have shown that when the party with the bargaining power has some taste for honesty, and reciprocates good behavior, we can achieve outcomes that come close to the first best, with cooperation decreasing as we reach the end of the relationship. It is unrealistic to suppose that we can modify behavior so that individuals are invariably and disinterestedly honest. But these results show that perfect integrity is not necessary. Rather, if one can instill a small taste for trustworthiness, one can design contracts that yield outcomes that are close to the first best. There are many situations where malfeasance works against such trust and leads to lower economic performance. My proposal is that despite the inevitability of such foibles, a promising avenue for enhancing economic performance in organizations lies in the potential for small changes in values, such as a taste for trustworthiness, to lead to large incremental changes in performance.

REFERENCES

Arrow, K. J. (1974). *The Limits of Organization*, New York: W. W. Norton & Co.

Binmore, K. (1994). *Game theory and the social contract. Volume 1. Playing fair*, Cambridge and London: MIT Press.

Carmichael, H. L. and W. B. MacLeod (2003). "Caring About Sunk Costs: A Behavioral Solution to Hold-Up Problems with Small Stakes." *Journal of Law, Economics and Organization.* Spring **19** (1), 106–118.

Ensminger, J. (2004). "Market Integration and Fairness: Evidence from Ultimatum, Dictator, and Public Goods Experiments in East Africa." in J. Henrich, R. Boyd, S. Bowles, C. Camerer, E. Fehr, and H. Gintis, eds., *Cooperation, Reciprocity and Punishment: Experiments in 15 Small-Scale Societies*, London: Oxford University Press, pp. 356–381.

Falk, A., D. Huffman, and W. B. MacLeod (2007). "New Evidence on Employment Protection, Bonus Pay, and Labor Market Performance." Technical Report, IZA.

Farnsworth, E. A. (1999). *Contracts, 3nd edition*, Boston: Little, Brown and Company.

Fehr, E. and K. M. Schmidt (1999). “A Theory of Fairness, Competition, and Cooperation.” *Quarterly Journal of Economics.* **114** (3), 817–68.

——— , S. Gächter, and G. Kirchsteiger (1997). “Reciprocity as a contract enforcement device: Experimental evidence.” *Econometrica.* July **65** (4), 833–860.

Gibbons, R. (1997). “Incentives and Careers in Organizations.” in D. M. Kreps and K. F. Wallis, eds., *Advances in Economics and Econometrics: Theory and Applications*, Cambridge, UK: Cambridge University Press, pp. 1–37.

Gneezy, U. and A. Rustichini (2000). “A fine is a price.” *Journal of Legal Studies.* Jan **29** (1), 1–17.

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

——— and B. Holmström (1987). “The Theory of Contracts.” in T. Bewley, ed., *Advances in Economic Theory: Fifth World Congress.*, Cambridge, U.K.: Cambridge University Press, pp. 71–155.

Henrich, J., R. Boyd, S. Bowles, S. Camerer, E. Fehr, H. Gintis, and R. McElreach (2001). “In Search of Homo Economicus: Behavioral Experiments in 15 Small-Scale Societies.” *American Economic Review.* **91** (2), 73–78.

Hoffman, E., K. McCabe, and V. L. Smith (1996). “Social Distance and Other Regarding Behavior in Dictator Games.” *American Economic Review.* June **86** (3), 1–8.

Kahneman, D. and A. Tversky (1979). “Prospect Theory: An Analysis of Decisions Under Risk.” *Econometrica.* **47**, 262–91.

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

Rabin, M. (1993). “Incorporating Fairness Into Game Theory and Economics.” *American Economics Review.* December **83** (5), 1281–1302.

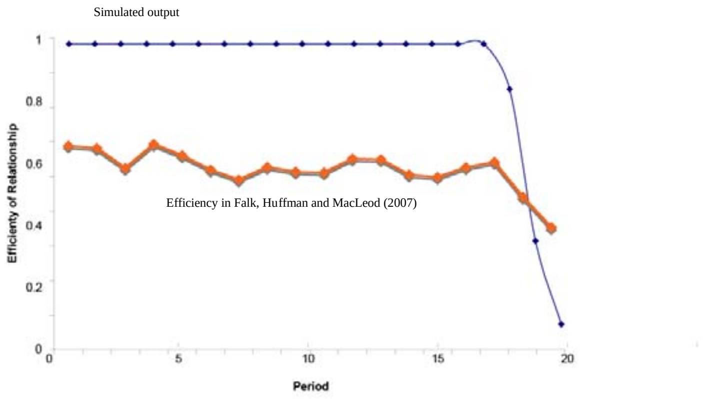

Figure 1: Productive Efficiency as a Function of Time

Smith, A. (1759). *The theory of moral sentiments*, London: Printed for A. Millar, and A. Kincaid and J. Bell, in Edinburgh.

Williamson, O. E. (1975). *Markets and Hierarchies: Analysis and Antitrust Implications*, New York: The Free Press.

Notes

[1] Page 23, Arrow (1974).

[2] See the Encyclopaedia Britanica entry on "convenant".

[3] Flow terms are used to simplify the analysis when varying the frequency. The total price in period t is therefore δp_t, total cost δq_t, resulting is a value of $\delta v(q_t)$.