

Self-interest through agency: An alternative rationale for the principal-agent relationship

John Hamman, George Loewenstein, Roberto A. Weber

Department of Social and Decision Sciences
Carnegie Mellon University
Pittsburgh, PA 15213

March 21, 2007*

Abstract

Economic actors are usually assumed to enter into principal-agent relationships due to efficiency gains from comparative advantage. This research explores whether such relationships may also arise because they allow principals to obtain a selfish, inequitable outcome without behaving themselves in an explicitly selfish manner. An experiment is reported in which principals either decide repeatedly how much money to share with recipients or select an agent to make the decision on their behalf. As hypothesized, principals keep more and recipients receive less when the allocation decision is made through agents.

Keywords: fairness, principal agent contracts, experiments

* We thank participants at the 2006 ESA meetings (Atlanta) and BDRM meetings (Los Angeles) and the 2007 ASSA meetings (Chicago) for comments and suggestions. We also greatly appreciate the financial support of the Russell Sage Foundation.

1. Introduction

In the standard economic analysis of the principal-agent relationship, principals are assumed to contract with agents because they confer efficiencies, either due to skill and expertise, or a lower opportunity cost of time or effort (see Hart and Holmstrom, 1987, for a review of the principal-agent literature). For example, agents can be used to delegate tasks for which the principal has a high opportunity cost, or to make decisions for the principal on matters for which the agent has greater expertise. The central focus of the literature on the principal-agent relationship has been on how to design monitoring and/or incentive schemes that enable these advantages to be realized despite the fact that agents typically face different incentives than the principals that employ them.

This paper examines whether the principal-agent relationship might in some cases serve a function unrelated to efficiency. Specifically, a principal may hire an agent to take self-interested or immoral actions that the principal is reluctant to take more directly. The principal may feel more detached, and hence less responsible, for such an action if so-delegated, while the agent may feel that he or she was “just carrying out orders” or merely fulfilling the requirement of an employment contract. Through the use of agents, therefore, accountability for morally questionable behavior can become vertically diffused with no individual taking responsibility.

While this function of agency has not been previously investigated in the economics literature, it is commonplace in popular accounts of behavior in domains as diverse as politics, business, war, and everyday social interaction. For example, companies are often accused of strategically outsourcing production and other functions to outside firms that act less ethically or socially responsibly than the company would act itself – for example, by treating workers less

generously or polluting the environment.¹ Executives and stockholders of the outsourcing company may turn a blind eye toward the actions of the outside firms, remaining deliberately uninformed, or at least pretending to be so. To give a more concrete example, one explanation for the Enron disaster lies in top executives' and board members' willingness to tacitly encourage the company's CFO and external auditors to engage in financial malfeasance on their behalf, while limiting their own personal involvement in such matters (Eisenberg, 2002; McLean and Elkind, 2003). Similarly, the steady rise in private military contractors over the last few decades – which was prevalent in previous centuries as well – is often linked to these entities' willingness to engage in warfare beyond what is permissible or politically feasible for public military forces, thus limiting public officials' accountability and responsibility for casualties or heinous acts (see, for instance, Thomson, 1995; Singer, 2007).²

In this paper, we test whether actions taken through agents are indeed likely to be more self-interested than actions one must take directly. We report results from an experiment in which principals, over 12 repeated rounds, face a decision of how much money to share with a passive and anonymous recipient, an experimental setup known as the “dictator game” (Forsythe et al., 1994). In the *baseline* condition, principals make the decision themselves, as is usually the case in experiments. In the *agent* and *agent/choice* conditions, principals select agents to make the decision on their behalf. After the agent chooses, the principal finds out what the agent chose and must decide whether to remain with the agent or select a different one. The agents are rewarded only for being selected.

¹ For example, in response to complaints about exploitative labor practices in independent factories in developing nations producing their products, Nike executives often claimed to have limited control over the actions of these outside firms. Moreover, firm managers who act unethically often justify such behavior through their role as agents for shareholders. For example, in discussing exploitative labor practices, Radin and Calkins (2006) note that one possible reason for their existence is because “Employees and managers view themselves as mere agents, and they act according to what they perceive are the values and interests of the principals” (p. 266).

² For example, much of the interrogation at Abu Ghraib prison in Iraq was conducted or overseen by independent contractors (Merle and McCarthy, 2004).

Our main prediction is that principals will share more money with recipients when they give directly than when they hire agents to allocate the money on their behalf. More precisely, we predict that principals, who would share significant amounts if choosing for themselves, will replace agents who do so on their behalf and will instead seek out agents who share little or nothing with the recipient.

In addition to investigating the impact of agency relationships on generosity, the experiment also examines what happens when principals are given the opportunity to decide whether to act individually or through agents. In rounds 9 – 12 of the agent/choice condition, principals are given the opportunity to choose whether to give directly to the recipient or select agents as in rounds 1 – 8. If principals value fair outcomes per se, they should choose to make the allocation decision themselves, thus ensuring that recipients receive a higher amount. And, if they view the actions taken by agents as fair, then they should be willing to take those actions themselves. However, a second prediction is that principals will continue to use agents who make selfish decisions on their behalf rather than making decisions themselves.

Finally, we also administer questionnaires at the end of the experiment in which we elicit several attitudes, including the degree of responsibility felt for the recipients' fate. We predict that subjects acting individually in the baseline condition will report feeling responsible, but that neither principals nor agents in the other conditions will report responsibility.

Existing Literature

Extensive research in economics suggests that people not only maximize their payoffs, narrowly construed, but display a strong preference for fair outcomes. (Kahneman, Knetsch, and Thaler, 1986; and see Camerer, 2003 for a review). Research on the dictator game shows that

even under conditions of complete anonymity, people are willing to share some amount of their allocation with an unknown recipient (Hoffman et al, 1994). Such results have led to the development of several theoretical models of fairness, which assume that people derive some utility from behaving fairly or some disutility from unequal outcomes (Rabin, 1993; Fehr and Schmidt, 1999; Bolton, 1991; Bolton and Ockenfels, 2000; Andreoni and Miller, 2002; Charness and Rabin, 2002; Loewenstein, Thompson & Bazerman, 1989).

Dana, Weber and Kuang (2004) present a series of experiments that call into question the notion of a stable “fairness preference.” Using several variants of a binary dictator experiment to examine what they call “moral wriggle room,” they find that agents act more selfishly when a mechanism is introduced that allows them to avoid direct responsibility for behaving selfishly.

One treatment from Dana et al (2004) is especially closely related to the current study. That treatment grouped two “dictators” with a single recipient. Both dictators simultaneously chose between a fair and selfish outcome, and if either chose the fair outcome then that result would occur. Thus, while either dictator could implement the fair outcomes, responsibility for the selfish and unfair outcome was shared. A substantially greater fraction of dictators behaved selfishly in this situation than when they were simply paired with a recipient and made an equivalent unilateral choice.

One possible explanation for these and other related findings, lies in models in which fair behavior is not motivated by a preference for fairness per se, but is based on the desire to maintain a positive self image, manifested through one’s actions (see, for instance, Prelec and Bodner, 2003; Benabou and Tirole, 2002; Konow, 2000; Murnighan, Oesch, and Pillutla, 2001). In these models, one avoids taking unfair actions because of what doing so implies about one’s self. However, if the objectionable actions do not reflect directly on one’s self, then it is much

easier to act in a self-interested fashion. Thus, much fair behavior occurs when, even though people may prefer to behave selfishly, they are constrained from doing so because of what self-interested and unfair actions would imply about themselves.

One way in which people can avoid making inferences about their own behavior is when others are involved in a decision, as in Dana et al (2004)'s experiment. Studies of "diffusion of responsibility," a concept first introduced by psychologists in the 1960s, show that individuals feel significantly less responsibility for actions undertaken by a group than they feel from their own individual decisions (e.g. Darley and Latane, 1968). This may arise from what Darley and his colleagues refer to as the bystander apathy effect (Latane and Darley, 1968, Darley, 2002). They found, for instance, that individuals are less likely to aid someone in distress if there are others who could do the same, to the extent that less aid is rendered when more people are present. In line with the diffusion of responsibility literature, principals in our current study may not feel responsible for being selfish if someone else makes the decisions (i.e. the "group" being the principal and the agent).

In contrast to the diffusion of responsibility literature, and to Dana et al (2004, study 3), which examine what could be called "horizontal" diffusion of responsibility across people in similar situations, the current experiment examines a setting in which one individual is subordinate to another, as in the standard principal-agent relationship. In such a setup, if agents act selfishly, principals can plausibly feel that they did not make the decision themselves, despite their position of power. Agents, likewise, may feel that they were simply responding to competitive pressures or "just following orders." The idea that each party would use different and self-serving principles to judge their own moral responsibility is consistent with the above research on self-serving judgments of fairness, which shows that people tend to judge the

fairness of their own behavior by drawing selectively upon norms of fairness those that are consistent with their own interests (see Babcock & Loewenstein, 1997, for a review).³

2. Experimental Design

Subjects were recruited by email from a list of graduate and undergraduate students interested in participating in experiments at Carnegie Mellon and the University of Pittsburgh. Either 12 (baseline) or 15 (agent, agent/choice) subjects participated in each session, each of whom received a \$7 show-up fee plus any additional earnings from the experiment. Each session lasted 30–45 minutes. All sessions were conducted at the Pittsburgh Experimental Economics Laboratory (PEEL) at the University of Pittsburgh. The experiment was conducted via computer interface, and was programmed using the software z-Tree (Fischbacher, 1999).

The experiment consisted of three conditions, which are illustrated in Table 1. In each condition, 12 subjects played the dictator game – 6 as dictators and 6 as recipients – for 12 rounds. The agent and agent/choice conditions also included 3 agents, denoted by the letter C. Subjects received their role assignment by randomly drawing a card with the letter A or B, as well as C in the agent and agent/choice conditions.

After role assignment, instructions detailing the rules of the game were given out and read aloud. Subjects were told that in each round dictators and recipients (roles A and B) would be randomly paired. They would know the ID number, but not the identity, of the person with

³ Principals also make use of agents in bilateral bargaining to strategically commit to advantageous settlements (Schelling, 1960; Bagwell, 1995; van Damme and Hurkens, 1997) and in the laboratory (Fershtman, Judd, and Kalai, 1991; Van Huyck, Battalio, and Walters, 1995; Croson, 1998). Prior studies have examined the use of agents for this purpose in bilateral bargaining situations (Katz, 1991; Blount, 1995; Schotter, Zheng, and Snyder, 2000; Fershtman and Gneezy, 2001). To the best of our knowledge, however, no studies have examined the psychological motivations for using agents discussed here.

whom they were paired. In each round, dictators received an allocation of \$10 to be divided between themselves and their randomly matched recipients.

In the baseline condition, dictators made the allocation decision themselves by specifying an amount they wished to share with the recipient. The allocation was shown to the paired recipient, and the experiment proceeded to the next round.

In the agent and agent/choice conditions, allocation decisions were made by one of the three agents. In the first round, dictators were randomly matched with an agent to ensure that all agents began with an equal number of decisions. In subsequent rounds (2-12), each dictator selected one of the three agents to make the allocation decision on the dictator's behalf, by clicking on that agent's number. In all rounds, each agent made the allocation decisions sequentially for each dictator, meaning that an agent selected by multiple dictators could allocate different amounts on their behalf. Any agents not selected in a round saw a waiting screen while the other agents made their allocation. After agents made their allocation decisions, each of these was shown to the corresponding dictator and recipient.

The agent/choice condition was identical to the agent condition through round 8, after which subjects received new instructions. For rounds 9–12, dictators were given the option of making the allocation decision on their own or continuing to select an agent.

In all conditions, once allocations had been made in each round, subjects saw the ID number(s) of other subjects with which they were directly involved with (including that of the agent, if applicable), and the amount of money transferred from the dictator to the recipient. Subjects recorded all results on their record sheets before continuing.

All subjects were told they would receive payment in cash at the conclusion of the experiment. Dictators and recipients were paid for one randomly chosen round, drawn at the end

of the experiment (plus the \$7 participation fee). Each agent began the experiment with \$5 and received payment in each round based on how many dictators selected him or her to make the allocation decision. More precisely, agents earnings in a round were based on the following payoff function:

$$\pi_i = -\$0.60 + \$0.30n_i, \quad (1)$$

where π_i is measured in dollars, and n_i represents the number of dictators (out of a possible 6) choosing agent i . This payoff function reflects a situation where agents face a fixed cost of providing service (\$0.60), zero marginal cost, and a constant marginal revenue (\$0.30). Agents, thus break even when they are “hired” by one-third, or 2, of the dictators.⁴

Notice that an agent’s earnings in a round are independent of her allocation decision, and vary only based on the number of dictators choosing her. This flat incentive structure was chosen for two reasons. First, it keeps the total payment to agents in each round fixed at \$0 and makes the net surplus available to agents was equal to the per capita surplus available to dictators and recipients (\$5 per person), thus eliminating net surplus and fairness considerations in agents’ compensation. Second, it avoids directly rewarding an agent for behaving in a fashion that maximizes the dictator's payoff, but instead allows the incentives for agents to arise more subtly, by dictators’ decisions of whether to remain with the previous agent or switch across rounds. Therefore, this is a “conservative” design; since the agents must tacitly infer what decisions the dictators are likely to reward, and allows the dictators to communicate these preferences only via their switching decisions.

⁴ We presented the payoff function to subjects in a slightly different format. We informed them that each agent would receive \$0.20 for every dictator who selected that agent, and would lose \$0.10 for every dictator that did not. When there are 6 dictators (as in the agent condition and the first 8 rounds of the agent/choice condition), this is identical to the payoff function above. However, in the final four rounds of the agent/choice condition, this alternative format allows us to easily keep the zero-sum nature of the payoffs by having them only apply to dictators not opting to make the choice for themselves.

Following the game, but before learning the random round that would count for payment, all subjects filled out a brief questionnaire with questions addressing feelings of fairness, responsibility, and enjoyment of participating in the session (each item provided 5 response choices ranging from -2: “Strongly Disagree” to +2: “Strongly Agree,” with a “Neutral” option), along with basic demographic questions.

Once all completed questionnaires were collected, participants were informed of which round had been randomly chosen to count for payment and shown their earnings. They then filled out their payment receipts and received payment.

3. Results

Giving to Recipients

Figure 1 shows the median amount given to recipients in each round for all three treatments.⁵ Recall that in rounds 1-8, both agent and agent/choice conditions were identical. By round 5, median amount shared drops to \$0.10 or less in both agent conditions and remains there, while the baseline condition median stays around \$1.50. This difference strongly supports the prediction that giving decreases when allocations are made through an agent.

Table 2 shows the average amounts shared in each round, presenting both the pooled agent conditions (for rounds 1-8) and also each condition separately. Note that in rounds 5 through 11, in both agent conditions the average amount shared is always lower than in the baseline – usually by over \$1, and always by at least \$0.50. Therefore, the agent treatments clearly decrease sharing relative to the baseline.

⁵ Two subjects – both dictators – in one session of baseline condition were found to have previously participated in the experiment, and so their results have been dropped from the analysis. Since dictators’ choices in the baseline condition are completely independent – dictators never find out what other dictators chose – we retain the choices of the remaining 4 dictators in that session. Therefore, the total number of dictators in the baseline condition is 40.

To compare overall sharing, we compute the average amount shared by each dictator over the first 8 rounds, when the two agent treatments were identical. In the baseline, dictators on average shared \$2.26 over the first 8 rounds, while principals in the agent conditions shared only \$1.70, which is significantly lower in a t-test ($t_{110} = 2.53$, $p = 0.01$) and a Mann-Whitney (rank-sum) test ($z = 3.34$, $p = 0.001$). While the difference of roughly \$0.50 is not large, this is because sharing in the agent conditions begins at high levels, similar to those in the baseline. For example, if we compare only rounds 1-4, we observe slightly higher sharing in the agent conditions than in the baseline (baseline: \$2.20; agent: \$2.29) and this difference is statistically insignificant ($t_{110} = 0.30$; $z = 0.10$). However, if we compare average sharing in rounds 5-8, we observe much larger and statistically significant differences (baseline: \$2.32; agent: \$1.11; $t_{110} = 5.08$, $p < 0.001$; $z = 5.20$, $p < 0.001$).

Figure 2 presents the distributions of amounts shared across conditions in rounds 5-8. Note that while at least \$1 is shared in the baseline 62 percent of the time, this generosity is much less frequent in the two agent treatments, where it occurs only 30 percent of the time. Similarly, the proportion frequency with which at least \$5 is shared is much higher in the baseline (28 percent) than in the agent conditions (11 percent).

Clearly, the agent treatments decrease sharing relative to the baseline, as we predicted, though this difference takes several rounds to emerge. Next, we explore whether this difference is due, as we predicted, to principals switching away from agents who shared large amounts.

Switching Agents

Beginning in round 2, all dictators in the agent conditions could choose among the three agents in their session, including the possibility of keeping the same agent as in the previous

round. Switching behavior provides insight into the preferences of dictators. If dictators value fair outcomes, as one might infer from the baseline, they should stick with agents who share amounts comparable to those shared in the baseline condition. However, if dictators prefer a selfish outcome to be implemented by the agent, they should search for agents who give very little or nothing at all to recipients.

Figure 3 shows the distribution of amount shared in periods 1 through 7 of the agent conditions, and the proportions of principals who remained with or switched away from an agent in the subsequent rounds (2 through 8). The solid portion of each bar represents the proportion of principals, who had that amount shared on their behalf, who chose to remain with the same agent in the following round. The shaded portion of each bar represents the proportion of principals who switched to another agent.

The figure supports the hypothesis that dictators switch away from agents who share sizable amounts in favor of those who share very little. For example, agents shared between \$0 and \$0.90 on a principal's behalf 288 times (the leftmost bar). In 260 of these instances (90 percent), the dictator chose the same agent again in the next round, and this proportion is slightly higher (93 percent) when the agent shared nothing. However, the frequency with which the principal retained the same agent decreases substantially with the amount shared. For example, of the 66 cases in which agents shared \$5, the principal only retained that agent 39 percent of the time.

Table 3 estimates with a fixed effects logistic regression the degree of switching in round t as a function of the amount given in period $t - 1$. The data again consist of rounds 2 through 8 from the pooled agent conditions. The large, highly significant coefficient on the variable

“GivenLast” supports the claim that the more an agent shares with recipients, the more likely it is that dictator will switch to a different agent in the subsequent round.

Session-level Heterogeneity

While Table 2 reveals that average sharing in the agent condition is lower than in the baseline, a closer look reveals a large amount of heterogeneity between two sets of agent sessions. Figure 4 shows that, of the seven sessions, two produced considerably higher mean sharing than the others. This appears to be mainly driven by fair and unpredictable agents in those two sessions. For example, even in round 1 – when agents were randomly-assigned and their choices were independent, the six agents in those two sessions shared significantly more than agents in the other five sessions (\$4.25 vs. \$1.28, $t_{40} = 4.06$, $p < 0.001$; $z = 3.64$, $p < 0.001$). Thus, unfair agents in these sessions were initially harder to find. Moreover, agents in these two sessions were also more unpredictable. Among agents selected in consecutive rounds by a principal, the proportion sharing the same amount again was 22 percent in these two sessions, and 75 percent in the other five sessions.

As a result of agents’ greater fairness and unpredictability, principals were much less likely to retain their previous agent in these two sessions (58 percent of the time) than in the other five (86 percent). Moreover, the relationship between amount shared by an agent and the principals’ subsequent decision to retain the agent was similar to what we observed previously. Principals retain agents who shared \$5 only 38 percent of the time (14 of 37). But in the 13 instances in which a principal found an agent willing to share \$0, the principal selected the same agent again 10 times (77 percent). However, of these 10 times, the agent only shared \$0 in the next round 5 of the times.

Thus, in these two sessions, principals were unable to find willing “accomplices,” given the small pool of agents from which to choose, and the result was a significantly higher frequency of sharing than in the other agent sessions. This suggests that a property of our experiment – the small number of potential agents – may increase sharing in the agent conditions, and that we might obtain even less sharing if we increase the number of available agents..

Questionnaire Responses

Table 4 displays mean responses, by role, for three questions taken from the questionnaire. These three questions address perceived fairness, responsibility for the recipients’ outcomes, and overall enjoyment of the experiment. As predicted, dictators in the agent conditions felt significantly less responsible for the amount given to recipients than baseline condition dictators. Agent condition dictators also felt they behaved more fairly than did baseline condition dictators, despite the more disparate earnings between themselves and recipients. Agent condition recipients disagree, reporting more unfair treatment than recipients report in the baseline condition. All dictators report similar levels of enjoyment, but recipients report less enjoyment in the agent conditions.

Table 5 shows a series of ordered probit regressions with questionnaire responses as the dependent variable. The independent variables are average earnings across all rounds (recall that subjects completed the questionnaire before finding out which round counted for payment) and a binary variable indicating condition (0 = baseline, 1 = agent or agent/choice). Dictators’ perceptions of the fairness of their actions is heavily influenced by how much they earned, but also by the condition – dictators in the agent condition believed they behaved more fairly, even

after controlling for average earnings. Conversely, recipients believed that they were treated significantly less fairly in the agent condition, even after controlling for their average earnings. Dictators' feelings of responsibility are strongly influenced by the experimental condition, with dictators in the agent conditions viewing themselves as significantly less responsible than those in the baseline. Finally, with respect to enjoyment, we find no effect of either treatment or earnings for dictators, but recipients' enjoyment was related to their average earnings.

Final Rounds in the agent/choice condition

Recall that, in the agent/choice condition, dictators could opt to make the allocation decision themselves in the final four rounds. We hypothesized that dictators who had first made use of agents, when subsequently given the opportunity to play themselves, would continue to select agents to make the decision on their behalf. This was only partially supported by the data. As shown in Table 6, roughly forty percent of subjects continued to select an agent, while the remaining sixty percent opted to make the decision themselves. Surprisingly, however, those dictators making their own decisions behaved almost identically to agents, sharing very little (roughly \$0.10 on average). While the dictators opting to make their own allocation decisions are essentially playing the last four rounds in the baseline condition, they share significantly less than what is shared in the baseline.

There are at least two plausible explanation for the results in Table 6. First, the 40 percent of subjects opting for an agent might be those who would feel compelled to share in a condition such as the baseline, without really wanting to, and might therefore be relying on the agents to implement a selfish allocation, as we hypothesized. The remaining 60 percent might be dictators who would not share very much anyway and therefore see little reason to rely on an

agent to do so.⁶ This is related to the findings of Lazear, Malmendier and Weber (2006), who show that dictators most likely to share large amounts are also among the most likely to avoid dictator games, thus keeping their money, whenever possible. Another potential explanation is that the previous 8 rounds, in which agents made decisions and shared very little, possibly desensitized principals to the recipients' fate or established a norm of very little sharing, which then carries over to the final four rounds. Therefore, aside from decreasing sharing directly, the agent treatment might have the perverse effect of extinguishing the intrinsic motivation to share by dictators. Our data does not allow us to distinguish between these two explanations

4. Conclusion

The research presented here suggests that principal-agent relationships can serve functions beyond those usually attributed to them in the economics literature. Specifically, principal-agent relationships can lead to an avoidance of responsibility for antisocial, immoral actions, and thus enhance the ability of principals to take such actions in pursuit of self-interest.

Principal agent relationships serve this function through a subtle interplay of psychological factors. Principals do not feel that they are behaving immorally because they are not directly taking immoral actions; they are simply hiring agents. Agents do not feel that they are behaving immorally because they feel they have no choice: "I was just following orders," "I had to do it if I wanted to make a living," or "if I didn't do it he would have hired someone else who would have done it," are emblematic of the types of phrases one often hears, after the fact, in investigations of atrocities, business fraud, and other forms of immoral behavior. In the real

⁶ In fact, these dictators, who want the selfish outcome and have no problem implementing it themselves, may dislike the uncertainty associated with allowing an agent to implement it. This uncertainty avoidance seems especially important given the fact that many agents gave away large amounts of money – including the entire \$10 – in the final round of both the agent and agent/choice conditions.

world, these psychological factors often interact with others that reinforce the diffusion of responsibility – for example, self-serving biases and willful disregard of information (see, e.g., Karlsson, Loewenstein & Seppi, 2005; Haisley & Weber, 2006).

Like the findings by Dana, Weber, and Kuang (2007) discussed earlier, the current study calls into question the idea, implicit in most models of social preference, that people have stable preferences for fair or equitable outcomes. If individuals know their outcome cannot be directly attributable, whether by others or by themselves, to their actions, people tend to become more selfish. One way in which this occurs is when the actual decision to behave selfishly or pro-socially is delegated to an agent.

A limitation of our study is that the study is quite conservative in two ways: (1) principals learned only about the behavior of the agents they hired, so they could only locate agents who would support their own interests through a slow search process, and (2) there were only three agents, which introduced a high likelihood that principals would be unable to find *any* agent who would act selfishly on their behalf. In fact, in two of the agent sessions we observed principals repeatedly seeking out agents who would act selfishly, but unable to find them. In the real world, there are likely to be many more agents to choose from in any situation, and agents are likely to have a well established reputation. Both of these factors would tend to strengthen the effects observed in our study.

Another limitation is that we only explore one kind of principal-agent relationship with one kind of contract (whereby agents receive a fixed payment for their service). In naturally-occurring relationships, there are many kinds of possible contracts, with varying incentives for the agent, and the selection of one kind of contract is endogenous to the relationship. Therefore, it would be interesting to explore how behavior and outcomes change under different kinds of

contracts – for example, contracts in which the principal establishes incentives for the agent (cf. Fershtman and Gneezy, 2001). Moreover, this would help distinguish whether it is the use of agents *per se*, uncertainty about what agents will do, competition among agents, or some combination of these factors that reduces the principals’ perceived responsibility for the outcome.

One prediction of the current research is that, in competitive situations in which individuals or firms could benefit from immoral conduct, one should be likely to observe the development of “agency” institutions through which self-interest can be delegated. Many organizational researchers have hinted at this, and some have taken a very negative view on such institutions from a sociological point of view. In a passage representative of this line of thinking, Korten (2001) writes that:

Behind its carefully crafted public relations image and the many fine, ethical people it may employ, the body of a corporation is its corporate charter, a legal document, and money its blood. At its core it is an alien entity with one goal: to reproduce money to nourish and replicate itself. Individuals are dispensable. (p. 73-74)

An even more general, and perhaps less pessimistic, take-away message from the current research is that agents may not always serve the function that they are ostensibly hired to perform. Our study suggests that agents may sometimes be used to avoid taking responsibility for one's actions, but other functions are possible. Thus, for example, the more thrifty member of a couple may propose to hire a financial advisor, not truly to provide investment advice, but to convey to the spendthrift partner that he or she needs to cut back on spending. Similarly, a restaurant may provide diners with the opportunity to delegate their dietary intake to someone likely to enhance flavor at the expense of a healthy diet, while the customer maintains some degree of uncertainty regarding the exact ingredients going into the food.

At a practical level, this paper points to potential institutional arrangements that could reduce immoral behavior in hierarchical situations. For example, any mechanism that forces principals to confront their own role in the outcome should, according to our analysis, tend to decrease immoral behavior. Again, a new treatment could address this issue: Once the agents make their decisions, all dictators could be forced to see the decision and have the option of over-riding it or certifying that they favor the action being implemented. Such a mechanism could potentially reintroduce the social pressure to behave altruistically. Indeed, a mechanism somewhat along these lines was part of the Sarbanes-Oxley bill that was enacted in the wake of Enron and other corporate scandals; it requires CEOs of companies to personally vouch for the accuracy of the books, thus undermining the common excuse of those at the top that they were unaware of, or otherwise not responsible for, malfeasance that occurred “in the ranks.”

References

- Andreoni, J., and Miller, J. (2002) Giving According to Garp: An Experimental Study of Rationality and Altruism, *Econometrica*, 70, 737-753
- Babcock, L., & Loewenstein, G. (1997) Explaining bargaining impasse: the role of self-serving biases. *Journal of Economic Perspectives*. 11, 109-126.
- Bagwell, K. (1995) Commitment and Observability in Games, *Games and Economic Behavior*, 8, 271 – 280.
- Benabou, R., and Tirole, J. (2002) Self-Confidence and Personal Motivation, *Quarterly Journal of Economics*, 117 (3), 871-915.
- Blount, S. (1995) When Social Outcomes Aren't Fair: The Effect of Causal Attributions on Preferences, *Organizational Behavior and Human Decision Processes*, 63, 131-144.
- Bolton, G. E. (1991) A Comparative Model of Bargaining: Theory and Evidence, *American Economic Review*, 81, 1096-1136.
- Bolton, G.E, and Ockenfels, A. (2000) ERC: A Theory of Equity, Reciprocity and Competition, *American Economic Review*, 90(1), 166-193.
- Camerer, C. (2003) Behavioral Game Theory: Experiments on Strategic Interaction, Princeton: Princeton University Press, 2003.
- Charness, G., and Rabin, M. (2002) Understanding Social Preferences with Simple Tests, *Quarterly Journal of Economics*, 117(3), 817-869.
- Coleman, J.S. (1974) Power and the structure of society. New York: Norton.
- Croson, R. (1998) Theories of Commitment, Altruism and Reciprocity: Evidence from Linear Public Goods Games, Discussion Paper, Wharton School, University of Pennsylvania.
- Dana, J., Cain, D., Dawes, R. (2005) What you don't Know Won't Hurt me: Costly (but quiet) Exit in a Dictator Game, *Organizational Behavior and Human Decision Processes*, 100(2), 193-201.
- Dana, J., Weber, R. A., Kuang, J. (2004) Exploiting moral wriggle room: Behavior inconsistent with a preference for fair outcomes, in press, *Economic Theory*.
- Darley, J.M. (2002) Crowded Minds: The Implicit Bystander Effect, *Journal of Personality and Social Psychology*, 83, 843-853.
- Darley, J.M., and Latane, B. (1968) Bystander intervention in emergencies: Diffusion of responsibility. *Journal of Personality and Social Psychology*, 8, 377-383.

- Eisenberg, D. (2002). "Is Ken Lay a Victim Too?" *Time Magazine*, Sunday, February 3.
- Fehr, E., and Schmidt, K. (1999) A Theory of Fairness, Competition, and Cooperation, *Quarterly Journal of Economics*, 114, 817-868.
- Fershtman, C. and Gneezy, U. (2001) Strategic Delegation: An experiment, *RAND Journal of Economics*, 32, 352-368.
- Fershtman, C., Judd, K., and Kalai, E. (1991) Observable Contracts: Strategic Delegation and Cooperation, *International Economic Review*, 32, 551-559.
- Fischbacher, U. (1999) z-Tree - Zurich Toolbox for Readymade Economic Experiments - Experimenter's Manual, Working Paper Nr. 21, Institute for Empirical Research in Economics, University of Zurich.
- Forsythe, R., Horowitz, J., Savin, N.E., and Sefton, M. (1994) Fairness in Simple Bargaining Experiments, *Games and Economic Behavior*, 6, 347 - 369.
- Haisley, E., and Weber, R. A. (2006). "Favorable Impressions of Ambiguity and Unstable Preferences for Fairness." Working paper.
- Hart, O. D. and Holmstrom, B. (1987) The theory of contracts, *Advances in Economic Theory*, Fifth World Congress, edited by T. Bewley. New York: Cambridge University Press.
- Hoffman, E. McCabe, K., Shachat, K., Smith, V. (1994) Preferences, property rights and anonymity in bargaining games, *Games and Economic Behavior*, 7, 346-380.
- Kahneman, D., Knetch, J., Thaler, R.H. (1986) Fairness and the assumptions of economics, *Journal of Business*, 59, 285-300.
- Karlsson, Niklas, Loewenstein, George F. and Seppi, Duane J., "The 'Ostrich Effect': Selective Attention to Information about Investments" (May 5, 2005). Available at SSRN: <http://ssrn.com/abstract=772125>
- Katz, M.L. (1991) Game-Playing Agents: Unobservable Contracts as Precommitments, *RAND Journal of Economics*, 22, 307-328.
- Konow, J. (2000) Fair Shares: Accountability and Cognitive Dissonance in Allocation Decisions, *American Economic Review*, 90 (4), 1072-1091.
- Korten, D. (2001) When corporations rule the world. Second edition. Kumarian Press, Bloomfield, Connecticut, USA.
- Latane, B. and Darley, J.M. (1968) Group inhibition of bystander intervention, *Journal of Personality and Social Psychology*, 10, 215-221.

- Loewenstein, G., Thompson, L., & Bazerman, M. (1989) Social utility and decision making in interpersonal contexts. *Journal of Personality and Social Psychology*, 57, 426-441.
- Lazear, E., Malmendier, U., Weber, R. (2005) Sorting in Experiments. Stanford working paper.
- McLean, B. and T. Elkind. (2003). *The Smartest Guys in the Room: The Amazing Rise and Scandalous Fall of Enron*. Penguin: New York.
- Merle, R. and E. McCarthy. (2004). "6 Employees from CACI International, Titan Referred for Prosecution." *Washington Post*, Thursday, August 26; page A18.
- Murnighan, K., Oesch, J.M., and Pillutla, M. (2001) Player Types and Self Impression Management in Dictator Games: Two Experiments, *Games and Economic Behavior*, 37(2), 388-414.
- Prelec, D. and Bodner, R. (2003) "Self-signaling and self-control," *Time and Decision*, G. Loewenstein, D. Read, and R.F. Baumeister (eds.) Russell Sage Press, New York, 2003.
- Rabin, M. (1993) Incorporating Fairness into Game Theory and Economics, *American Economic Review*, 83, 1281-1302.
- Radin, T. J. and M. Calkins. (2006). "The struggle against sweatshops: Moving towards responsible global business. *Journal of Business Ethics*, 66: 261-272.
- Samuelson, W. and Zeckhauser, R. (1988) Status Quo Bias in Decision Making, *Journal of Risk and Uncertainty*, 1, 7-59.
- Schelling, T. (1960) *The Strategy of Conflict*. Harvard University Press, Cambridge, Mass.
- Schotter, A., Zheng, W., and Snyder, B. (2000) Bargaining through agents: An experimental study of delegation and commitment, *Games and Economic Behavior*, 30, 248-292.
- van Damme, E. and Hurkens, S. (1997) Games with Imperfectly Observable Commitment, *Games and Economic Behavior*, 21, 282-308.
- Van Huyck, J., Battalio, R., Walters, M. (1995) Commitment versus Discretion in the Peasant-Dictator Game, *Games and Economic Behavior*, 10, 143-17

Table 1: Description of experiment by condition

	Baseline	Agent	Agent/Choice
Rounds 1 – 8	Self	Agent	Agent
Rounds 9 – 12	Self	Agent	Agent/Self option
Sessions	7	7	5
N (dictators, agents, recipients)	84 (42, n/a, 42)	105 (42, 21, 42)	75 (30, 15, 30)

Table 2: Mean amount given comparisons

Round	Baseline (n = 40)	Pooled agent conditions (n = 72)	Comparison with baseline		Agent (n = 42)	Comparison with baseline		Agent / choice (n = 30)	Comparison with baseline	
			(a)	(b)		(a)	(b)		(a)	(b)
1	2.59	2.59	0.27	0.00	2.13	0.72	0.84	3.23	1.45	1.08
2	1.87	2.83	1.24	1.72*	3.16	1.00	1.94*	2.37	1.15	0.95
3	2.31	2.02	1.06	0.58	2.35	0.72	0.07	1.54	1.14	1.48
4	2.05	1.74	1.42	0.60	1.57	1.74*	0.86	1.96	0.56	0.14
5	2.40	1.45	1.62	1.90*	1.34	1.84*	1.92*	1.60	0.83	1.29
6	2.38	1.10	2.21**	3.01***	1.24	2.08**	2.23**	0.90	1.66*	2.90***
7	2.26	0.91	2.81***	3.70***	0.91	2.79***	3.20***	0.90	1.91*	2.78***
8	2.25	0.98	2.90***	3.03***	1.18	2.42**	2.23**	0.71	2.54***	3.01***
9	2.23				1.73	1.64*	0.96	0.13	4.51***	5.60***
10	2.12				1.38	2.02**	1.50	0.18	4.44***	4.84***
11	1.95				1.04	2.13**	2.01**	0.08	4.12***	4.53***
12	1.81				2.97	0.43	1.62	0.59	2.55***	2.68***

a: Mann-Whitney rank-sum (z); b: t-test; both two-tailed

Table 3: Fixed effects logistic regression of switching in round t based on amount given in round $t - 1$. Data pooled from both agent conditions for rounds 1 – 8. 17 subjects (119 observations) dropped due to all positive or all negative outcomes. Dependent variable: Switching in t .

	Coef.	Std. Err.	z	p	N
Givenlast	.473	.065	7.27	0.000	385

Table 4: Questionnaire responses with pooled A-Ab data

Question	Dictators / Principals		Agents	Recipients		Sig. Diff. (t-test)
	(A) Baseline (n=40)	(B) Agent (pooled) (n=72)	(C) Agent (pooled) (n=36)	(D) Baseline (n=42)	(E) Agent (pooled) (n=72)	
I {made fair decisions / was treated fairly} in this experiment.	0.05	0.32	0.06	-0.62	-1.26	A/B: p<.001; D/E: p=.01
I feel responsible for how much the [recipients] were allocated.	0.73	-0.13	0.69	--	--	A/B: p<.001
I enjoyed this experiment.	0.85	0.69	-0.14	0.07	-0.51	D/E: p<.05

Table 5: Ordered probit regressions of questionnaire responses

	Dictators / Principals			Recipients	
	Fair Decisions	Responsibility	Enjoyment	Fair Treatment	Enjoyment
Condition (Agent)	0.363* (0.200)	-0.719*** (0.200)	-0.288 (0.302)	-0.495** (0.226)	-0.332 (0.218)
Average Earnings	-0.265*** (0.088)	-0.025 (0.078)	0.100 (0.065)	0.202** (0.104)	0.245** (0.101)
N	112	112	112	114	114

Standard errors in parentheses (clustered by session)

* - $p < .10$; ** = $p < .05$; *** - $p < .01$

Table 6: Results of rounds 9 – 12 in agent/choice condition

Round	Allocation Decision		Average Amount Shared		
	Agent	Self	Agent	Self	Baseline
9	13	17	0.17***	0.10***	2.23
10	12	18	0.08***	0.19***	2.12
11	12	18	0.08***	0.08***	1.95
12	10	20	1.51	0.13***	1.81

Differences with baseline: * - $p < .10$; ** = $p < .05$; *** - $p < .01$

Figure 1: Median giving across rounds in each condition

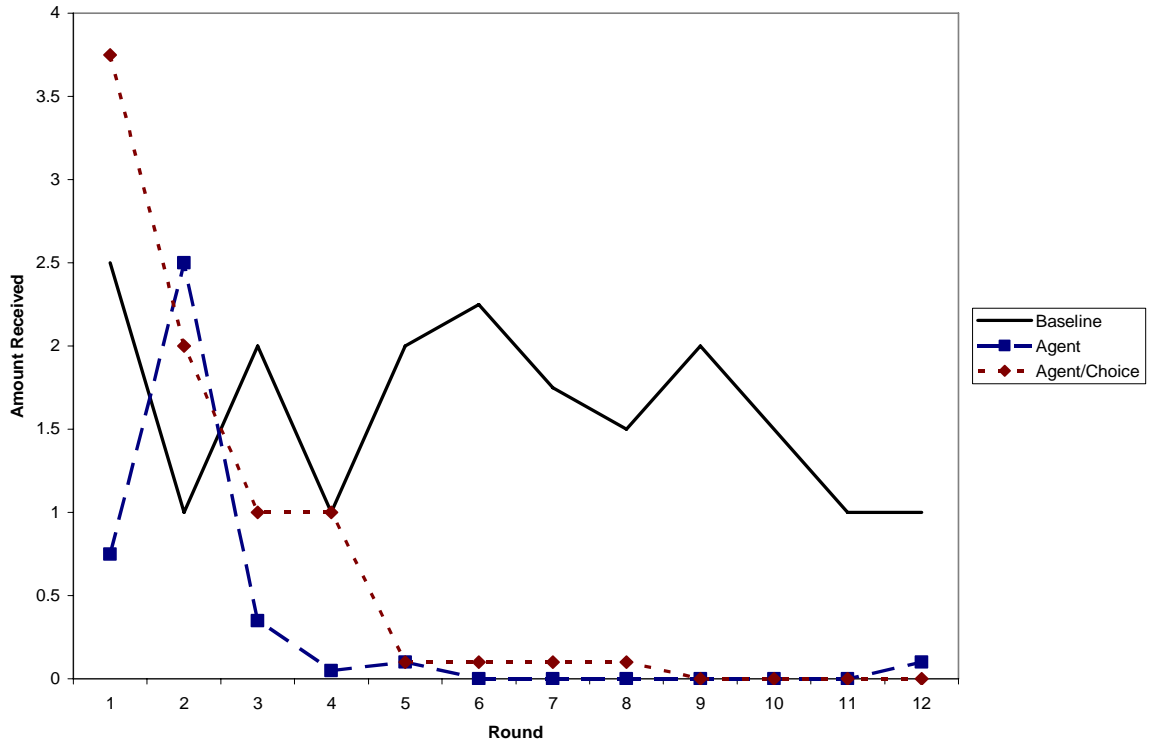


Figure 2: Distribution of amount given by condition, Rounds 5 – 8

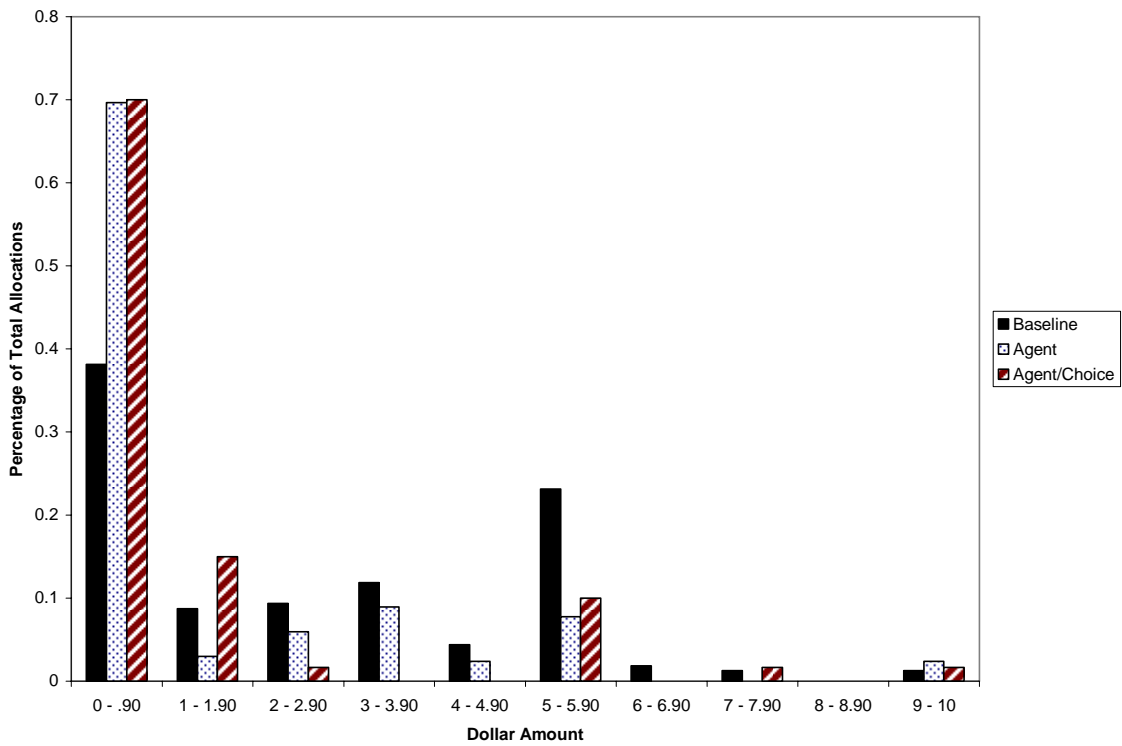


Figure 3: Switching in round t based on amount given in round $t-1$. Data pooled from two agent conditions, using rounds 2 – 8.

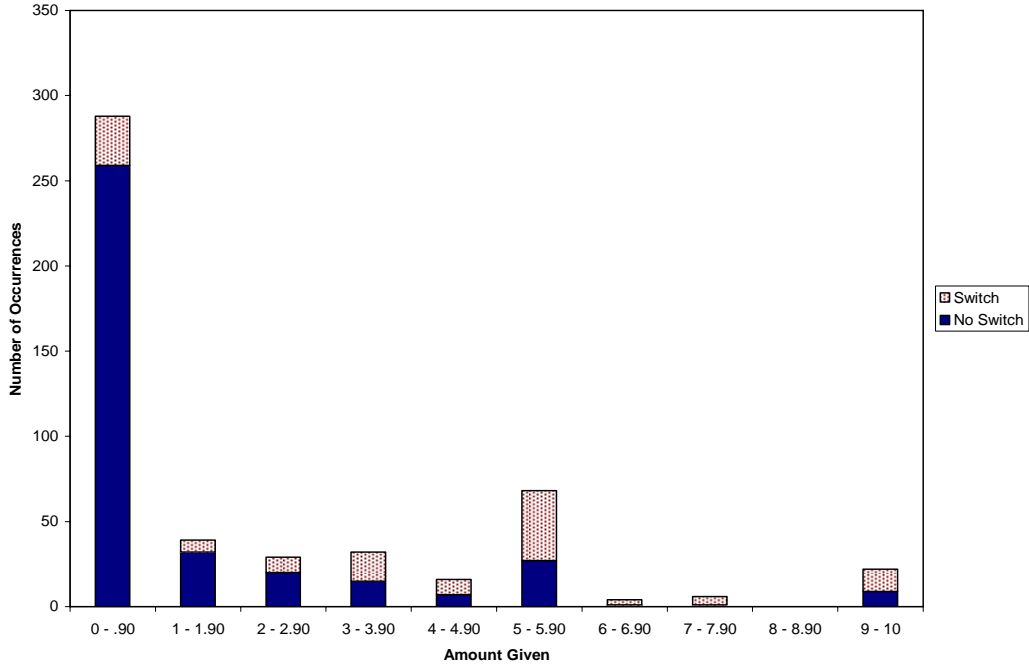


Figure 4: Mean giving in agent condition, by session

