Intermediation Reduces Punishment and Reward*

Lucas C. Coffman†

October 1, 2009

Abstract

Most economic theories of punishment assume the moral decision-maker has in mind a model of fairness - based on outcomes, intentions and procedures - and punishes deviations from fairness. In contrast, this paper shows that moral decision-making is better modeled as an unreasoned, intuitive reaction, correlated but not tethered to the punisher's ideals of right and wrong. Specifically, the experiments in this paper demonstrate that misbehavior is punished substantially less when its consequences do not directly follow; moral judgment is immoderately focused on the immediate outcomes of an action. Moreover, this narrow judgment is shown to greatly reduce the effectiveness of punishment and reward, which allows anti-social behavior to go unsanctioned and pro-social behavior unrewarded. In Study 1, in the laboratory, the action of keeping money at the expense of a poorer player is punished less when done through an intermediary, even when the intermediary is not involved in selecting the outcome. In the game, the intermediary can only make the poorest player worse off, so current fairness models would predict an increase in punishment, the opposite of these results. Many social preference papers find a discontinuity between positive and negative actions, so in Study 2, I test the robustness of the psychology behind the main result from Study 1 in a charity/reward domain. In Study 2, a framed field experiment, reward of charitable behavior (donating mosquito nets) decreases when the saliency of an intermediary (the charity) is increased. Together the results show that moral decision-making is not well-predicted solely by normative ideals. As a result punishment and reward are unreliable alone in enforcing pro-social equilibria.

JEL classification: C92, D63

Keywords: moral decision-making, fairness, intermediation, punishment, experiments

*I owe much to Alvin Roth, Max Bazerman, Sendhil Mullainathan, Muriel Niederle, Gary Charness, and David Laibson for guidance. I would also like to thank Michael Faye, Judd Kessler, Stephen Leider, Bazerman Lab and conference participants at ESA Tucson 2007, Pasadena 2008, and Lyon 2008 for comments.

†lcoffman@hbs.edu; Harvard University, 25 Harvard Way, Baker Library 433, Boston MA 02163

1 Introduction

> *"Each society learns to live with a certain amount of... mis-behavior; but... society must be able to marshal from within itself forces which will make as many of the faltering actors as possible revert to the behavior required for its proper functioning."*
>
> - Albert O. Hirschman (1970)

In 2005, Merck sold the patent for a cancer drug to a small pharmaceutical company, Ovation, that immediately raised the price charged to patients by a factor of 10.[1] Firms often worry about negative reactions to price-gouging (e.g. Kahneman, Knetsch & Thaler 1986), but was Merck not equally concerned about consumer outrage in response to selling to a firm that might raise prices so dramatically?[2] Outside of "passing the buck" or diffusing responsibility, is Merck avoiding blame simply by not doing the deed themselves? That is, what if Merck was known to be solely responsible for and solely beneficial from the price increase? Could Merck possibly avoid some loss of goodwill merely because the price increase is not perceived to be a direct consequence of their actions?

If moral decision-making depends on something as inconsequential as directness, this would be evidence that punishment is not just a response to fairness as economists see it, based on outcomes, intentions, and procedures. If punishment were a perfect response to these variables, it could reliable in encouraging pro-social behavior, and there is much evidence that this is the case. Punishment can effectively regulate individuals (Fehr & Fischbacher 2004, Fehr & Gächter 2000) as well as companies (Dyck & Zingales 2004, Martin & Thomas 2005). Increased pro-social behavior means that punishment can lead to greater social welfare for the group (Axelrod 1986).[3] However, the Moral Psychology literature provides evidence that punishment and moral judgment are not merely responses to the objective wrongness of an action (Frederick & Kahneman 2002). Rather, moral preferences are at least partially derived from an intuitive or emotional reaction (Greene & Haidt 2002, Haidt 2001), so perhaps we should expect punishment to be not only about what is *optimal* for the group but also about what *feels* right or wrong about what just happened. Our results support this notion that punishment is the product of an unreasoned, intuitive process, correlated but not tethered to aligning an individual's incentives with those of the group. In some cases, as in this paper, moral intuition is actually contrary to aligning individual incentives with those of the group. In these cases, as we will see, pro-social equilibria will not be supported.

[1] Berenson, Alex. 2006. "A Cancer Drug's Price Rise is Cause for Concern." *New York Times*. March 12, Business section.

[2] The small company, Ovation, raised prices on three different drugs in 2005 by at least a factor of 10 shortly after purchasing them.

[3] Similarly, punishment can be sufficient for collusion amongst oligopolists (Stigler 1964) even with imperfect monitoring of each others' price and quantity decisions (Green & Porter 1984; Abreu, Pearce & Stacchetti 1986).

Further, the Moral Psychology literature provides evidence that blame increases the more directly the consequence follows from the action (Royzman & Baron 2002; Greene, Lindsell, Clarke, Nystrom, & Cohen, 2008). Paharia et al (2009) even show that subjects who read hypothetical scenarios rate firms as less immoral if the firms act antisocially through an intermediary actor. Such results are consistent with much behavior found in the marketplace but cannot be explained by economic models of fairness and punishment. This paper introduces these ideas from Moral Psychology to Economics.

This paper analyzes the unreasoned nature of moral decision-making in a moderately complex environment commonly found in market transactions - when there is a third party between an action and its consequence. This third party can be anything from Ovation separating Merck from the price increase, to any agent, lawyer, consultant, subcontractor, outsourced firm, third party vendor, or market mechanism that may disjoin a direct action-outcome link. In particular, this paper examines how perceived morality changes when intermediation only serves to sever this direct link between an action and its consequence. We do not study buck-passing or responsibility diffusion. We find that actors receive substantially less punishment or reward for a given action when they use an intermediary actor, even a passive one. It seems as though the proximity of the action to its consequence increases the perceived goodness or badness of the act.

The primary experimental design employed here (the *Intermediation Game*) allows one economic agent (the first mover) the option of playing a \$10 Dictator Game (DG) directly with an unendowed receiver or gives her the option to unilaterally "sell" the game to an intermediary who already has \$5. If the first mover sells the DG, she also decides the price that the intermediary must pay her for it. In the resulting DG, the intermediary must keep at least as much as he was forced to pay for the DG; thus, using the intermediary is equivalent to allowing him the opportunity to take a slice of the unendowed receiver's share. It follows that, holding constant the first mover's wealth, intermediation makes the poorest player (the receiver) weakly worse off. An unaffected observer then has the opportunity to punish the first mover. Holding fixed the first mover's profit, she is punished significantly less when these profits come indirectly (via intermediation) rather than directly (in the DG), making intermediation the profit-maximizing strategy. Moreover, subjects correctly believe that intermediation reduces equity greatly; thus, the main result cannot be explained by current outcome-based (Fehr & Schmidt 1999, Bolton & Ockenfels 2000) or intentions-based (Rabin 1993, Dufwenberg & Kirchsteiger 2004) models of fairness and runs counter to models based on both outcomes and intentions (Falk & Fischbacher 2004).

Perhaps most importantly, since outcome norms are not as strongly enforced with intermedia-

tion, socially preferable outcomes occur less frequently: The poorest player is much worse off, and total surplus decreases substantially. These results come from comparing the results from the *Intermediation Game* to a treatment without the availability of intermediation (the *No Intermediary* treatment, akin to the *Third Party Punishment Game*, Fehr & Fischbacher 2004). On average, the unendowed receivers in the *No Intermediary* treatment make 2.7 times more than their counterparts in the *Intermediation Game*. Further, since the poorest player is worse off in the *Intermediation Game*, punishment is higher (despite being tempered by the presence of an intermediary), which is Pareto-damaging. The non-fixed portion of subject payouts decreases 22.6% in treatments when an intermediary is available. Changing to an environment where moral judgment is impaired can greatly reduce both pro-social behavior and efficiency.

A follow-up study confirms the result holds in a case of charity and reward: When the saliency of intermediation is increased for charitable giving, rewards decrease. The experiment varied the saliency of the fact that charitable donations pass through an intermediary before reaching the recipient. While the increased saliency of an intermediary had no effect on giving, it had large, negative effects on rewards. Together, these results suggest that indirectness of behavior attenuates both reward or punishment, and thus perhaps that indirectness decreases the total perceived "goodness" or "badness" of the act.

It is known that responsibility can be diffused when a second party is present[4]. Charness (2000), Fershtman & Gneezy (2001), and Bartling & Fischbacher (2008) all demonstrate experimentally that the addition of a second responsible party may alleviate perceived responsibility of the first party. Fershtman & Gneezy show this is true even though the first mover is incentivizing the second party to act selfishly on her behalf and even when she has the choice of not using the second party at all. The authors provide the insight that perhaps the behavior they observed is "because it is not [done] directly by the proposer but by a third party," though they do not provide experimental support for this specific hypothesis. Bartling & Fischbacher show similar abated reciprocity even though the first mover is delegating the choice to a second party who has incentive to act selfishly on behalf of both of them. The experiment presented here is designed to allow for analysis of scenarios where the intermediary is completely blameless for the outcome. Even in such scenarios, punishments decrease for actions done through an intermediary. This result does not refute responsibility diffusion, but rather identifies a separate force acting in tandem, strengthening

[4]This paper follows the custom of assuming punishment, blame, responsibility and unfairness are positively correlated. As Falk, Fehr & Fischbacher (2005) conclude, "the desire to harm those who committed unfair acts, seems to be the most important motive behind fairness-driven sanctions." It must be noted, however, that though these are very similar, they are not identical. Cushman (2008) shows punishment is better predicted by outcomes while judgments without consequence are better predicted by the intentions of the actor.

the claims that the frequency of bad actions may increase when more people are involved, and that one may want to involve more people when taking a bad action.[5] The strategic implications of the current result, though, might differ from those associated with diffusion of responsibility. If a profit-maximizing agent can avoid punishment by adding an *unresponsible* second party, then she can clearly increase profits by doing so. In contrast, sharing responsibility may imply sharing profits, and the strategic implications for profit-maximization are unclear.

The result is also more than the first mover obfuscating her action. While it is true that using an intermediary can make finding evidence of the first mover's culpability more difficult, this will not be a factor in our experiment. Clearly, Merck could avoid blame if the public had no idea Merck ever even owned the drug or if it did, for how much Merck sold it. This paper investigates a much tougher test of the effect of intermediation on blameworthiness. Our experiment makes the entire history transparent.

We are interested in intermediation because of its ubiquity. Throughout, "intermediation" refers to any intentionally added layer between an actor and the outcome. Intermediations such as Merck's sale of its cancer drug are common. Many companies may face less backlash for outsourcing production to firms who cut costs with questionable labor practices or by allowing more pollution than they would face if they took the same actions themselves. Intermediation can be, and often is, less obvious than selling or outsourcing. Fehr, Hart & Zehnder (2008) show that experimental "suppliers" would shirk less often when paid a low price if the "buyer" defers price determination to a market process with a known low outcome rather than choosing a low price directly. Charness (2004) similarly shows reciprocity is lower when wage is determined by an external process. There are also myriad real world examples of firms having a staffing company or even an industrial psychologist lay off workers, hoping perhaps that the employees might harbor less ill will towards the firm if employers do not deliver the pink slip themselves. Intermediation is not restricted to firm and worker relationships - in divorce proceedings, it would be inappropriate if one called his soon-to-be ex wife a terrible mother (especially in front of one's children), but maybe this impropriety is attenuated if one has his lawyer say it on his behalf. Like all of these examples, the experiment within will consider human intermediaries, though perhaps the concept is more generalizable; giving a friend cash for her wedding might be considered uncouth in America, but if you give her the money through an online gift registry (by purchasing the item she has already selected), this might be viewed as more palatable.

[5]Hamman, Loewenstein & Weber (2008) show guilt might also decrease in such situations, and Charness & Dufwenberg (2006) suggest that guilt can be an important motivation.

2 Experimental Design - *The Intermediation Game*

The Intermediation Game has four players - first mover, intermediary, receiver, and punisher (See Appendix, "Experiment Instructions" for exact instructions as given to participants). The first mover owns a dictator game (henceforth "DG") worth \$10 to be played with the receiver. She has two options of what to do with the DG. She can play the game herself, or she can "sell" it to the intermediary (Whether the term "sell" mislead subjects is discussed in Section 4 in the analysis following Result 2). If she chooses to sell, she also chooses the price the intermediary must pay her for the DG. The first mover's strategy space is the same whether she is choosing how much to keep in the DG or the price for which she will sell it: {\$5, \$6, \$7, \$8, \$9, \$10}. If the DG is sold to the intermediary, she plays it with the same receiver, and she must keep at least as much as she was forced to pay; she cannot lose money. The intermediary has a \$5 endowment, equal to half the size of the DG (For a discussion of whether subjects' failure to include Player B's \$5 endowment in their fairness judgments drives the results, please see Appendix subsection 6.2). Thus, holding fixed the rents extracted by the first mover, using the intermediary weakly reduces equity and weakly makes the poorest player (the receiver) worse off.

Finally, the punisher can then reduce the payoff of the first mover (and only the first mover). The punisher may base her punishment decision on (i) the amount sent in the DG, (ii) whether the DG was sold, and (iii) for how much the DG was sold. The punisher can reduce the first mover's payout to any non-negative amount, and punishment is costless. Punishment decisions are elicited via the strategy method: The punisher gives a decision for all possible outcomes of the game the other three are playing, and her decision for the scenario that is realized is enacted[6]. The punisher is endowed with \$5. Thus, the intermediary and punisher both begin with \$5, and the first mover is deciding how to split \$10, so a 4-way, \$5 equal split is clearly feasible and can be guaranteed by the first mover if she does not sell the DG.

The intermediary has no say in the selling or pricing decisions, so the experiment avoids any complicity of the intermediary in the first mover's rent extraction. Punishment was costless in order to increase power. The task of the punisher is akin to the task of a judge. Judges are not assessed costs for penalties they mete out; the effect their decisions have on others are incentive for truth-telling and thoughtful decision-making. The data do not reveal any patterns of careless

[6]The strategy method may induce "cold" decision-making on the part of the punisher. In Paharia et al (2009), differences in moral judgments of direct versus indirect acts disappeared when the two scenarios were judged side-by-side; the similarity in the wrongness was too salient. Thus, we expect this design decision should attenuate any punishment differences we may observe. There is no reason to suspect it would drive or increase the main results.

decision-making, so this assumption seems valid.[7] Since the strategy method was utilized, to reduce the number of scenarios posed to the punisher, all transactions in the game were in multiples of $1 (Sell prices and DG allocations). This reduced the number of scenarios to 27. The punisher was thus asked for 27 punishment decisions, one at a time, in a random order. The order was reversed each subsequent session. The same random order was used in other treatments (as well as the reversal if there were two sessions). In each session, the game was repeated four times, with each subject playing each role once. They were informed that they were re-matched into new groups in each period (To see an analysis of order effects and within-subject consistency, see Appendix Section 6.1).[8] They were paid for their earnings in one randomly chosen period. This is to discourage thinking of the game as a four period meta-game but rather focus on maximizing each period. To minimize learning about others' behavior, no feedback was given until all four periods were completed.

After the four periods, but before feedback was given, subjects were asked for their beliefs of how much the intermediary kept for each transaction price. They were paid $0.50 if they were within $0.50 of the average of how intermediaries actually behaved in that session (and $0.50 if this scenario did not occur in this session)[9]. In the final session of the *Intermediation Game*, they were then asked to guess the beliefs of others. They were paid $0.50 if they were within $0.50 of the average guess of everyone in the room from the questions previously answered. After these third party, second order beliefs were obtained, feedback was given.

All experimental sessions were performed at the Computer Lab for Experimental Research (CLER) at Harvard Business School. All sessions consisted of 12, 16, 20, or 24 study participants recruited via email from the Boston area. All subjects were under the age of 30.[10] They were paid a $10 show up fee plus their earnings from the experiment, which averaged $4.57 for the game and $3.37 for the belief elicitations. The experiment was conducted at computer terminals using z-Tree 2.1.4 (Fischbacher, 1999). No session took longer than one hour.

[7] In pilot experiments, punishment was costly. There was less punishment overall, but the punishment dynamics described in section 4 were the same.

[8] Given sessions as small as 12 subjects, we could not eliminate being in the same group as another subject twice. Subjects were never in the same group as another subject in consecutive periods. When they were in the same group twice, the subject did not know whom or when.

[9] This incentivizes subjects to report the midpoint of the $1 interval they believe to be most likely for their session. Beliefs were elicited in this way in order to make it simple and understandable for the subject.

[10] In one session, the recruitment software experienced a bug and allowed subjects over 30 to sign up. These data are not included in the analysis. The main results do not change with their inclusion. In sum, over-30's punished more, also punished intermediation less, and exhibited more outcome bias. The over-30's were confused by the instructions and were unfamiliar with computers, and consequently they did not finish in one hour and were let go.

Control - *No Intermediary Treatment*

To evaluate the welfare effects due to the presence of intermediaries, a control experiment, with no intermediary, was also run. Thus the control was an instantiation of the *Third Party Punishment Game* (TP-DG, Fehr & Fischbacher 2004). Otherwise, everything was consistent with earlier sessions: punishment was costless, the punisher made $5, and the DG was worth $10. Throughout, this will be referred to as the *No Intermediary* treatment.

Two Punishments Treatment

To understand better how subjects allocate blame, we added a treatment identical to the *Intermediation Game* but allowing the punisher to punish *both* the first mover and the intermediary. The punishment mechanism does not change: Punishment is costless, and the punisher cannot reduce either player's payoff to negative amounts. She can punish both, one or neither. The amount she chooses to punish one does not affect her strategy space for the other punishment decision. The punishments are again elicited using the strategy method, and though the scenarios are presented one at a time, the punishment decision for the first mover and the intermediary are elicited simultaneously.

The Reflection Treatment

To encourage subjects to think about and see if they understood the game and particularly the strategies of other players, they were asked in this treatment to think about the game and write down their thoughts beforehand. This treatment was exactly the same as the *Intermediation Game* except subjects spent four minutes before the game writing on four blank sheets of paper for a minute apiece reflecting on four questions: (1) Why might the first mover sell the game? (2) Why might the first mover not sell the game? (3) What will happen if the first mover does not sell the game? (4) What will happen if the first mover sells the game? The questions are intentionally neutral and balanced across scenarios of selling and not selling. This treatment was to encourage the subjects to reflect about what is going on in the game and to test if the results hold for subjects who are clearly aware of the potentially dubious motives of other players. As Paharia et al (2008) suggest, moral judgments seem to be intuitive, and when subjects are forced to reflect on motives, judgments can change significantly.

3 Behavioral Predictions

There is a class of models that make clear predictions for the *Intermediation Game*. Interestingly, none will consistently predict less punishment for rents extracted through the intermediary. Moreover, the models with the best standing up until now in predicting behavior will predict that punishment will be *greater* for money made via intermediation. That is, the public might punish Merck *more* for selling the drug to an un-punishable intermediary.

The simplest way to separate the competing theories is to ask what punishment levels they predict when the payoff outcome is the same, but the path taken is different. For example, the outcome is the same if the first mover keeps \$X in the DG or if she sells the DG for \$X and the intermediary then keeps \$X, making zero profit. In both cases, the resulting payout is (\$X, \$5, \$10-\$X, \$5) for the first mover, the intermediary, the receiver, and the punisher respectively; recall the intermediary and the punisher are endowed with \$5 each. We will consider what punishment dynamics each class of models may predict for each path.

Outcome-based Models of Fairness: If the punishers have preferences based solely on the outcome, as in Fehr Schmidt (1999) or Bolton & Ockenfels (2000), then, punishments will be independent of whether the intermediary was used; punishment will be identical in the two cases raised above.

Intentions-based Models of Fairness: Rabin (1993) and Dufwenberg & Kirchsteiger (2004) define fairness through intent, and the relevant piece of intent for *The Intermediation Game* is beliefs:[11] If I believe you believe I will play a_i, and you play b_j, then I judge your fairness based on what profits would result from a_i and b_j regardless of what I actually play. Extending this framework for the *Intermediation Game*, beliefs take the form, "I believe the first mover believes the intermediary will keep \$Y when the DG is sold for \$X (with Y$\geq$X), so when the first mover sells the game for \$X, I will judge her for directly choosing the allocation (\$X, \$5+(\$Y-\$X), \$10-\$Y, \$5) regardless of the actual actions of the intermediary." Thus, in comparing this action (selling the DG for \$X) to not selling the DG and keeping \$X directly, the punisher judges the fairness of choosing (\$X, \$5, \$10-\$X, \$5) (no intermediation) versus (\$X, \$5+(\$Y-\$X), \$(10-Y), \$5) (intermediation). Notice that the two allocations differ only by a \$(Y-X) transfer from the receiver to the intermediary, a transfer from the poorest player to a player already endowed with the average pre-punishment payout (\$5). Rabin (1993) and Dufwenberg & Kirchsteiger (2004) do not define preferences over others' payouts (but rather focus on refining our understanding of intent). However, if a punisher

[11] The other piece of intent is the strategy space from which the judged action was chosen, but the strategy space is held constant in our game, so this will not be a part of the discussion

in their setup is indifferent over who receives the $(Y-X), punishment is expected to be identical whether the first mover keeps $X in the DG or via intermediation. If punishers are not indifferent about others' payoffs, these intentions-based models are also outcome-based and make the same predictions as the models discussed in the following subsection.

Intentions and Outcome-based Models of Fairness: Their experimental evidence leads Falk, Fehr & Fischbacher (2008) to conclude that the best models we currently have of predicting reciprocal behavior assume fairness is based not only on intentions but also on the outcome. Falk & Fischbacher (2004) provide the first such unifying framework. However, agents in their model only have preferences over outcomes relative to their own. We adapt this model to include preferences over third party outcomes using the general model of Charness & Rabin (2002). They claim, "[subjects] like to maximize the minimum payoff among players."[12] This seems a reasonable extension and is experimentally demonstrated by Charness & Rabin to be the salient factor in determining preference over others' payouts. Recall from the discussion in the *Intentions* subsection that intermediation weakly decreases the minimum payout, and more importantly, we will show subjects *believe* that the first mover *believes* that intermediation weakly hurts the poorest player. Thus, in this hybrid model, punishment is predicted to be weakly greater when the first mover uses the intermediary. More interestingly, this inequality will be strict if the punisher believes the first mover believes $Y-$X>0.

Diffusion of Responsibility: Charness (2000) and Bartling & Fischbacher (2008) both demonstrate that including a second, human agent who is at least partly responsible for the outcome at least partially absolves the initial agent. In Bartling & Fischbacher's model of this concept, an agent's responsibility is determined by her share of the total increase in the probability of the bad outcome occurring due to human agents; thus, if a second party increases the likelihood of the outcome, the principal's responsibility decreases for the same bad outcome, holding fixed her actions. However, in our game, even when the first mover sells the game, she has, at the very least, guaranteed the receiver cannot be paid more than $10 less the price the first mover demanded; thus, she is at least as responsible as she is when not selling the game and keeping the same amount. Further, by putting the DG in the hands of an unpunishable party, the probability of a worse outcome for the receiver weakly increases, so doing so, in fact, will make the first mover weakly *more* responsible than she would be had she not sold the game, and we would expect weakly more punishment. In addition, this model makes an even clearer prediction when the intermediary is perfectly altruistic;

[12]The prediction that will follow does not change if instead agents have a preference for equity over others' payoffs since intermediation also weakly damages equity. Other than indifference, equity and maximin are the leading candidates in the literature of preferences over third party payouts.

When the second agent does not transgress, responsibility is not diffused and punishment should not decrease. We should expect weakly more punishment for intermediating.

4 Results

The results in this section together tell an interesting story. Using an intermediary decreases punishment. We will show this is not due to diffusion of responsibility, nor confusion or lack of thought by the punisher. Rather, the results here will suggest punishment is not determined by reasoning and explicit beliefs of fairness. Punishment is guided by intuition, and that moral intuition judges indirect actions less harshly. This creates opportunity for actors to benefit from anti-social behavior while not facing large punishments. Selfish behavior and unequal outcomes will be common.

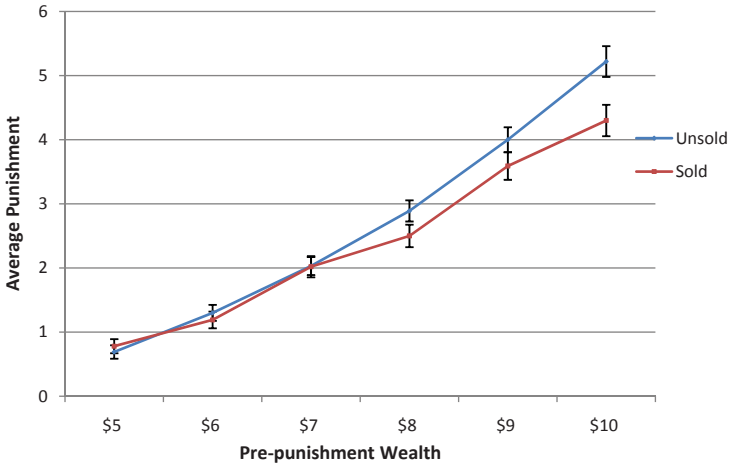
In *The Intermediation Game*, each punisher was asked to make a punishment decision for the same 27 scenarios (see Appendix, Table 6 for complete list of scenarios & average punishments) facilitating matched, nonparametric analysis of the data, which will be employed alongside OLS regression analyses. Appendix Table 7 provides averages of key variables for a general understanding of how the game was played. Unless otherwise noted in this section, reported results are from *The Intermediation Game* treatment.

81% of subjects punish at least once, and when they do punish, it is economically substantial, almost equal to average profit and fully 58% of the first mover's pre-punishment wealth (See Table 8 in the Appendix for more details on what may have determined the size of punishments). Figure 1 also shows just how large average punishments were in the experiment, with an average punishment over $5 for one scenario. Also note that 52% of subjects use the intermediary, so it is not an irrelevant alternative. Result 1 may help explain why.

Result 1 - Punishment significantly decreases when rents are extracted through intermediation rather than directly.

Figure 1 and Table 1 show that when the first mover keeps more than $7, directness is punished more harshly: Average punishment is greater when the first mover extracts rents directly, and many subjects punish in this direction - At these profit levels, at least 2.5 times as many subjects punish direct actions more harshly than subjects who punish indirect actions more harshly. Using nonparametric tests - both a Wilcoxon signed rank test and a Fisher Pitman matched pair permutation test - the punishment distributions are shown to be significantly different for each of

Figure 1: **Average Punishments**



these high profit levels of the first mover (See the last two columns of Table 1). In other words, when punishment (and misbehavior) is high, direct actions are frequently and significantly punished more.

Table 1: Directness is Punished More Harshly[a]

	Avg. Punishment		No. Subjects Who Punish		p-values	
Profit of First Mover	DG Unsold	DG Sold	Unsold More	Sold More	Matched Pair, Signed Rank[b]	Permutation Test[c]
\$10	\$5.22	\$4.30	24	7	<0.01	0.01
\$9	\$4.00	\$3.52	15	6	0.05	0.07
\$8	\$2.89	\$2.48	20	8	0.02	0.07
\$7	\$2.03	\$1.98	16	12	0.58	1.00
\$6	\$1.30	\$1.41	11	6	0.22	0.27
\$5	\$0.69	\$0.78	2	1	0.55	0.50

[a]To hold outcomes constant, only scenarios where intermediary makes no money included
[b]Matched Pair Signed Rank test based on null hypothesis that punishment distributions are identical for direct and indirect action. Pairs matched by subject.
[c]Permutation p-values based on 200,000 simulations per test. P-values do not change at the hundredths level when re-run. Pairs matched by subject. P-values are from two-tailed tests.

The strongest case in support of Result 1 is in the first row of Table 1: When the first mover keeps *all \$10* for herself, she is punished less when she does so indirectly. When sold the DG for \$10, the intermediary is not even making a decision; she had to pay \$10, so she has to keep \$10 in the DG. The first mover has effectively chosen the final outcome - (\$10, \$5, \$0). All that is different is that the intermediary has to press a button, having no choice other than to click on "Keep \$10". More than any other pair of scenarios, the intention and responsibility of the first

mover is transparently equivalent, yet punishment drops 18% when the intermediary is forced to press a button.

Result 1 is also confirmed in an OLS regression approach. Table 2 regresses punishment on the various monetary outcomes and paths taken ("Sold" is a dummy for whether the DG was "sold" to the intermediary). According to the point estimate in the first row in Model I of Table 2, for every dollar the first mover keeps over $5 playing the Dictator Game herself, punishment increases 90 cents. The coefficient on the interaction term in the second row is the linear estimate of how much this punishment changes when she sells the DG to the intermediary: She is punished 26 cents less for every dollar she takes by "selling" the DG, a 29% reduction. We will discuss model II later in this section and models III-V after Result 2.

Table 2: OLS Regressions

Dependent Variable = Punishment ($)					
	The Intermediation Game		Two Punishments Treatment		
	All Scenarios	Intermediary Makes $0	All	Never Diffuse	No Diffusion
	(I)	(II)	(III)	(IV)	(V)
	N=64	N=64	N=36	N=24	N=36
First Mover's Profit ($)	0.90***	0.90***	0.68***	0.67***	0.68***
over $5	(0.09)	(0.09)	(0.14)	(0.16)	(0.30)
First Mover's Profit	-0.26***	-0.18***	-0.19**	-0.21**	-0.30***
over $5 *Sold	(0.06)	(0.06)	(0.09)	(0.08)	(0.08)
Intermediary's Profit	0.05**	–	0.01	0.00	-0.07
	(0.02)		(0.01)	(0.00)	(0.05)
Sold Dummy	0.31***	0.16*	0.32	0.26***	0.20
	(0.08)	(0.09)	(0.20)	(0.08)	(0.12)
Period	-0.28	-0.29	0.32**	0.26	0.17
	(0.25)	(0.27)	(0.15)	(0.19)	(0.16)
Constant	0.43**	1.88	-1.00**	-1.05*	-0.67
	(0.21)	(1.67)	(0.39)	(0.52)	(0.40)
Session F.E.	Included	Included	Included	Included	Included

Standard Errors Clustered at Subject Level
***, **, * indicates p<0.01, 0.05, 0.1 respectively

Though this approach imposes linear relationships, it has some nice features. First, it gives an idea of the magnitudes. Second, it allows for inclusion of scenarios where the intermediary makes positive profits in the *Intermediation Game*, which is more difficult in pairwise nonparametric tests. This is important since punishers exhibit "outcome bias": Punishment increases slightly, the more the intermediary keeps in the DG. This can be seen on the coefficient on "Intermediary's Profit" in Table 2. Though only a point estimate, the reported magnitude of the outcome bias, however, is quite small ($0.05), less than one fifth the expected decrease in punishment for each dollar made when intermediating ($0.26). Moreover, if the game is sold for a price near $10, there is little profit

for the intermediary to make, and thus little opportunity for outcome bias to affect punishment.

These tests also speak to the behavioral predictions from Section 3. In section 3, we elicited predictions from leading classes of models for what punishments would be for scenarios which differ only in the path taken to the outcome. Table 1 makes just such a comparison. Recall, none of the predictions produced by the fairness models in Section 3 are consistent with punishment being greater for direct actions. However, Table 1 shows that subjects more frequently punish direct actions more harshly (Table 9 in the appendix furthers this point). Further, Model II in Table 2 re-runs the OLS specification in Model I, this time only including scenarios where the intermediary made \$0 profit. Note that the coefficient on the interaction term in the second row of Model II is also significantly less than 0; thus, even when the intermediary is *completely blameless*, her inclusion reduces punishment.[13] No model discussed in Section 3 is consistent with this result.

Result 2 - Blame is not just being shifted or diffused; it is simply decreasing.

In addition to considering scenarios where the intermediaries are *completely blameless*, we can further our case by looking at these scenarios only when zero blame is actually assigned to the intermediaries. Recall in the *Two Punishments* treatment, both the first mover and the intermediary could be punished. Thus we can re-run our analyses controlling for whether punishment was shifted or diffused to the intermediary, and this is driving result 1. Table 2 shows OLS regression output for such scenarios. All specifications only include data from scenarios where the intermediary made zero profit. Model IV controls for punishment-shifting at the subject level; it only includes data from punishers who *never* punished the intermediary in any of these scenarios. Model V controls for punishment-shifting at the punishment level; it only includes punishments of the first mover if the punishment of the intermediary was \$0. In both cases, punishment still decreases when the intermediary is used: The coefficient on the first mover's profit interacted with a dummy for selling the game is significantly negative in both cases (Moreover, this result holds if we restrict the analysis to the first mover keeping \$9 or \$10; see section 6.2).

This table simultaneously responds to concerns that perhaps subjects were confused, particularly by the framing of the game. In general, the term "sell" implies mutual consent for the transaction from the buyer and seller. It is very reasonable then to wonder if the subjects are thus assigning some blame to the intermediary simply because the word "sell" tells them she too was responsible. Table 2 shows that subjects who did not diffuse punishment, and thus were not confused in this

[13]Note the point estimate of the punishment decrease is directionally (though not significantly) smaller in Model II compared to Model I, indicating that while not driving Result 1, responsibility diffusion might explain some of the decrease in punishment when using an intermediary (about 1/3 of the total decrease).

way, show similar punishment dynamics in the dimensions of interest.

Result 3 - The availability of intermediation makes the poorest player dramatically worse off and reduces total surplus.

When the first mover sells the DG, she is putting the DG in the hands of an unpunishable agent, who sends, on average, only \$0.21 to the receiver. This shifts money from the receiver to the intermediary, from the poorest player to a player with the average wealth level. As Table 3 indicates, the receiver earns a paltry \$1.63 average in the *Intermediation Game* versus \$4.50 in the *No Intermediary* treatment.[14] In the Third Party Punishment Game (similar to the *No Intermediary Treatment*), Fehr and Fischbacher (2004) conclude that punishment is successful in encouraging significantly more equitable outcomes. When an intermediary is added to the same game, this is far from the case.

Not only is the availability of an intermediary bad for the poorest player, it is Pareto-damaging. Even though punishment is less when the intermediary is used, it is still greater than zero (The sum of the first two coefficients in Table 2, \$0.64, which is the amount of punishment per dollar kept when selling the DG, is significantly greater than \$0). Since punishment is inefficient, this creates a loss in overall social welfare. Table 3 also shows that average profits decrease by 54 cents per subject compared to the *No Intermediary* profits. This is the average for all four players, so over \$2 is lost among the group. This result is particularly surprising since \$10 of group earnings is fixed (the endowments of the punisher and intermediary), so 22% of what *can* be lost is lost when an intermediary is present (which amounts to 11% of total possible surplus).

Table 3: Welfare and Equality

	Intermediation Game	**No Intermediary Treatment**	**Rank Sum p-value**
Receiver's Avg. Payoff	\$1.63	\$4.50	<0.001
% Who Receive \$5	22%	79%	
% Who Receive \$0	55%	0%	
Avg. Profit (All Roles)	\$4.31	\$4.85	0.01

Result 4 - Even when the DG is unsold in the Intermediation Game, *the receiver is worse off than had there not been an intermediary.*

Dictators who do not to sell the DG keep more than dictators in the *No Intermediation* treatment;

[14] Average profits reported are for all periods, not just the randomly chosen period for which the subjects were paid.

they only send \$3.13, significantly less than the \$4.50 sent in the *No Intermediary* treatment (Wilcoxon two-tailed rank sum p-value<0.01). This could be due to selection: Those who send the *most* in the *No Intermediary* treatment might be now selling the DG to the intermediary.[15] It could also be a changing norm: The same person will now choose to keep more because there is an intermediary present even if she does not use the intermediary. We cannot separate these stories in this experiment though either is an implication of how intermediation changes the game that is worthy of future investigation.

Result 5 - Subjects believe intermediation makes the poorest player substantially poorer, and they believe others share this belief.

Subjects correctly predict the extreme inequity that intermediation allows. Recall beliefs of how much the intermediary kept in the DG were elicited in an incentivized manner after the game was played but before feedback was given. Beliefs of how much the first mover keeps in the DG were elicited in the *No Intermediary* treatment since this eliminates selection of who sells the game.

Table 4 shows that subjects expect the intermediary to keep at least 67% of the remaining value of the DG, leaving as little as \$0.05 and up to \$1.70 for the receiver. In stark contrast, subjects believe the first mover will send the receiver about \$4 when she is not allowed to sell the game (recall these beliefs, in the first column of Table 4, were elicited in the *No Intermediary* treatment). Moreover, 46 of 64 subjects report that the intermediary will keep more than she was made to pay for the DG for *every* price, *and* that others believe this as well (See Appendix, Table 11 for OLS regressions of these 46 subjects). Wilcoxon rank sum tests confirm subjects believe less will be sent to the receiver when the game is sold regardless of the price ($p<0.01$, two-tailed test comparing the unsold belief distribution to the of the distribution of beliefs for any given price). Moreover, as the second row of Table 4 shows, they believe other subjects hold similar beliefs. Thus punishers believe using the intermediary reduces equity and makes the poorest player much worse off; they believe the first mover shares this belief, and yet they punish her less when she uses the intermediary. This is particularly striking because experimental subjects have been shown to have a strong maximin preference over payoffs (Charness & Rabin 2002). Punishing intermediation less runs opposite this preference.

Subjects believe the first mover believes intermediation reduces the minimum payoff (and equity). Thus, according to fairness models based on beliefs, intermediating should be deemed an antisocial

[15] This would imply that the subjects who would choose the most generous allocation directly - those most motivated by guilt aversion, altruism, or blame-avoidance - are now choosing to intermediate; thus, they believe intermediating avoids guilt, maximizes altruism, or avoids blame.

Table 4: Beliefs of How Much Will Be Kept in the DG

		Purchase Price				
	Unsold	**$5**	**$6**	**$7**	**$8**	**$9**
Expected Amount Kept in $10 DG	$5.90	$8.37	$8.97	$9.30	$9.67	$9.85
Beliefs of Others' Expectations of Amount Kept	$6.10	$8.30	$8.90	$9.45	$9.75	$9.95
Actual Amount Kept	$5.50 N=24	NA N=0	$10 N=2	$9.43 N=7	$9.75 N=8	$9.83 N=6

action and punished *strictly* more harshly. Thus Result 3 makes the predictions of existing models fit even less well in this context.

If not beliefs, what is driving subjects' behavior? One hypothesis is that some subjects have limited rationality: Pinning responsibility on the first mover in the case of intermediation takes more reasoning ("The receiver is poor because the intermediary kept this much *because* the first mover sold the game for this much," versus, "The receiver is poor because the first mover kept this much."). Thus it could plausibly be that this increased rationality requirement is driving Result 1. This *limited rationality* hypothesis was the motivation for the final design, *The Reflection Treatment*. Recall, subjects were given four neutral questions as primes for reflection before the game was played. 22 of the 24 subjects wrote something for all four questions. Interestingly, 15 also explicitly mentioned in at least one of their answers that the first mover will sell the game in order to avoid punishment.

Result 6 - Limited Rationality is not driving Result 1.

Table 5 shows the OLS regressions for this subject pool (Model I) using the same specification as in Table 2. Model II is the same specification ran only on the 15 subjects who explicitly mentioned the first mover's punishment-skirting intention. In both models, rents kept over $5 are punished less when the intermediary is used.

Thus even subjects who wrote down that first movers sell the game to try to avoid punishment, punish first movers less when they sell the game. They exhibit an awareness of the dubious intentions of a first mover who sells, but this does not predict their punishment. Coupling this with Result 5 (subjects believe intermediation makes the poorest player substantially poorer), it has been shown that subjects who fully understand the poor outcomes and dubious intentions involved in intermediation still punish it less; thus, limited rationality does not seem to explain Result 1. Potential reconciliations of all these results are discussed Section 6.

Table 5: OLS Regressions: Reflection Treatment

Dep. Var. = Punishment ($)		
	All Subjects	15 "Wise" Subjects
	(I)	(II)
First Mover's Profit ($)	1.14***	1.09***
over $5	(0.16)	(0.21)
First Mover's Profit	-0.33***	-0.41***
over $5 *Sold	(0.10)	(0.12)
Intermediary's Profit	0.09***	0.04
	(0.03)	(0.03)
Sold Dummy	0.10	0.39*
	(0.29)	(0.20)
Period	-0.42	-0.29
	(0.31)	(0.38)
Constant	1.31	1.25
	(0.97)	(1.31)
Standard Errors Clustered at Subject Level		
***, **, * indicates p<0.01, 0.05, 0.1 respectively		

5 Study 2 - Intermediation Reduces Reward

A follow-up study was conducted with two main objectives - (i) to test the generalizability of Result 1 by placing it in a new context and (ii) to attempt to understand the mechanism driving Result 1. The second experiment was a modified *Intermediation Game* with charitable behavior (real donations to a real charity) and rewards rather than antisocial behavior and punishment.

Subjects were undergraduates at Harvard University. All were recruited via email in November 2008. Each dormitory was randomized into one of two experiments and sent a link for that experiment. The website they visited then randomized them into a treatment, so randomization within each experiment was at the individual level[16].

Survey 1 Design

Subjects in the first experiment were told that 3 names would be drawn for every 100 survey respondents. The first name drawn would win $70 and have the opportunity to donate money to purchase mosquito nets for pregnant mothers in Busia, Kenya[17]. This opportunity was framed as either a direct gift to the pregnant mothers or as a donation to a charity (TamTam, www.tamtamafrica.org) who would then purchase nets for pregnant mothers (See Appendix, "First Person Donation Fram-

[16] Experiment was conducted using Qualtrics survey platform, www.qualtrics.com
[17] Mosquito nets are very important in this region to protect against malaria.

ing" for exact wording). It was made clear that in both cases, a donation of $3.50 would result in a pregnant mother in Busia receiving a mosquito net and that the money would be sent to TamTam, who would deliver the net. The framing only manipulated the saliency of passing the donation through the intermediary. Subjects reported the number of nets they would like to donate, up to 20. Subjects were assured they could receive a forwarded email from TamTam if they chose to donate. This could ease any suspicions the money would not be sent while maintaining the nice feature that giving here is not to receive approbation from the receiver. The sender is anonymous, so her only motivations are charity and self-signaling (e.g. Benabou Tirole (2006) and Grossman (2009)).

The second and third names drawn would play the two roles of a Dictator Game worth $48 with efficient giving. The amount sent was scaled up by a factor of three (to encourage a great number of nonzero gifts). Everyone made both a mosquito net donation decision and a dictator decision. The dictator decision always followed their donation decision. Each question was on a separate screen, and at any point in the experiment, they did not know what would be on the next screen. They were not made aware of any second experiment.

Survey 1 Results

Please see Tables 14 for confirmation that the randomization worked. Subjects were 97% non-seniors, 69% women, and 59% participate in the student charity organization (PBHA). This is not representative of the undergraduate population, so there was sorting; however, even though this may be expected to affect the amount of giving, we have no reason to expect this may interact with the treatments.

The framing manipulation had little to no effect on donations. The 94 subjects in the direct frame on average donated 10.7 nets while the 108 in the indirect frame donated 10.2. While their generosity is noteworthy (donating more than half of the $70 prize), the manipulation is not (Wlicoxon ranksum testing equal distributions: p=0.61). This result is in line with the self-serving bias literature. Performing an act with an intermediary results in ambiguity of who is responsible. Subjects will resolve this ambiguity in a self-serving way; thus, if responsibility is good, as it is here, they will take full credit, and we should not see any difference between the two conditions[18].

Subjects who donated at least one net directly gave directionally, but not significantly, more in

[18]Hamman et al 2008 show that, indeed, subjects are more likely to undertake *unethical* endeavors with an intermediary, in such cases resolving the ambiguity by convincing themselves the intermediary is responsible.

the subsequent DG than subjects who donated at least one net indirectly: $9.70 vs $7.85 (ranksum p-value=0.22). This question was added primarily to elicit beliefs in an incentivized way in the second survey; we do not hypothesize any difference in altruism between who may give in the two conditions.

Survey 2 Design

Survey 2 elicited rewards of charitable behavior of subjects in Survey 1 as well as beliefs of their altruism. After a subject gave consent to participate, they were randomized into either the *reward* experiment or the *altruism beliefs* experiment. In either experiment, they were then told they would be randomly matched with one subject from Survey 1. They were additionally told one piece of information about this person. They were (correctly) told either:

> *Direct Framing:* "This respondent was asked if he/she won $70, would he/she like to purchase mosquito nets (at $3.50 each) for pregnant women in Busia, Kenya (through TamTam www.tamtamafrica.org). This respondent said yes - Had his/her name been drawn he/she would have purchased mosquito nets for pregnant women in Busia, Kenya."
> OR
> *Indirect Framing:* "This respondent was asked if he/she won $70, would he/she like to donate money to Tam Tam (www.tamtamafrica.org). Money donated to TamTam enables them to purchase mosquito nets (at $3.50 each) for pregnant women in Busia, Kenya. This respondent said yes - Had his/her name been drawn he/she would have donated money to TamTam."

Subjects in the *reward* experiment were then asked to play a dictator game (worth $84) with efficient giving (again scaled up by a factor of 3) with the anonymous Survey 1 respondent about whom they just learned. They were told the Survey 1 respondent would receive a letter, along with any money sent, explaining that another respondent was given the opportunity to send them money based on their decision to donate nets. Thus, this was framed as an unexpected reward. They received the letter even in the case of zero money sent.

Subjects in the *altruism beliefs* experiment did not reward but rather were asked to guess how much this anonymous Survey 1 respondent sent in the DG she played following her donation

decision. They were paid \$50 if they were within \$5 of how much the respondent actually sent[19].

Survey 2 Results

Subjects rewarded direct-framed giving more than indirect-framed giving. Subjects sent \$36.4 to "direct" donors (thus they received \$109.2), but subjects only sent "indirect" donors an average of \$29 (thus they received \$87). A Wilcoxon rank sum test of a hypothesis that these reward distributions are equal gives marginal significance (p=0.099). The cumulative distributions of rewards shown in figure 2 further show the difference in rewards. Rewards to the direct donor first order stochastically dominate rewards to the indirect donor.

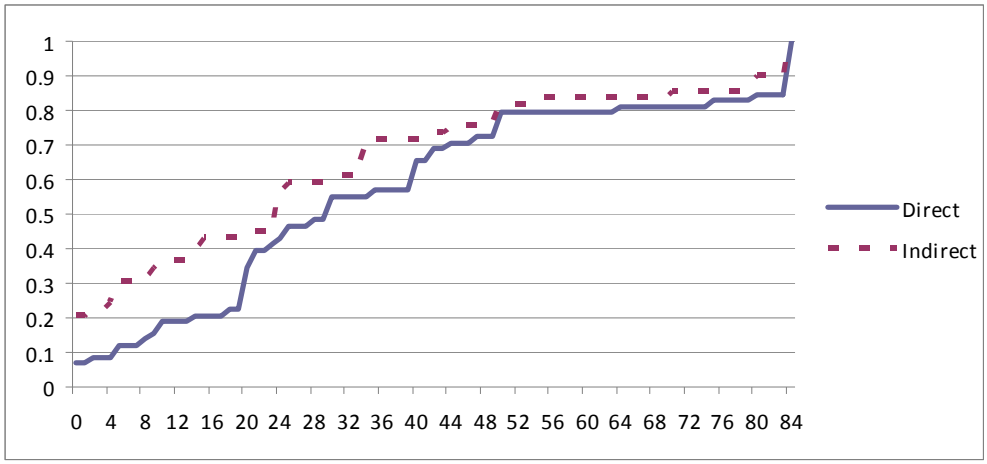


Figure 2: CDF's of Rewards to Donors, by Framing Manipulation.

This is preliminary evidence that the preferences driving Result 1 from *The Intermediation Game* are somewhat generalizable. We have changed the behavior from selfish to charitable, the first mover's motivation from intrinsic and extrinsic to intrinsic (the first mover knew of the punisher in *The Intermediation Game*, but not of the rewarder here), the environment from the lab to online, and from anonymous, undergraduate subject receivers to real charities and charitable cases. In this very different environment, a similar result develops: Intermediation reduces reward as it did punishment.

Finally, subjects believe subjects who donated nets indirectly donated *more* in the subsequent

[19]This elicits their belief of the midpoint of the \$10 interval containing the most Survey 1 donators in their distribution of beliefs. Beliefs were elicited this way in order to keep it simple and understandable for the subject.

DG. They guessed indirect donors sent \$14.1 compared to \$8.5 by direct donors. A Wilcoxon rank sum reports these distributions are significantly different, p=0.04. If the DG decision in Survey 1 is believed by subjects in Survey 2 to be independent of the decision to donate (it always came after), then this should be a good measure of baseline altruism. This question was added to test a hypothesis that Result 1 was being driven not by "punishing bad behavior" but rather "punishing bad people". As Levine (1998) argues, we might have more preference for the payouts of altruistic agents. This hypothesis is not supported in Survey 2; rewards are higher for the group with less expected altruism (direct donors).

Admittedly, the DG decision in Survey 1 may very well not be independent. There might be wealth effects, though wealth is only in expectation. Additionally, often after performing a good act, we feel licensed to be less moral in the next period (e.g. Cain, Loewenstein & Moore 2005). Whether subjects predict such inter-question dependencies, or others, is unknown to the author. Regardless, this result should only be regarded with these considerations.

6 Discussion

This paper investigates how the fairness of an action is judged when performed through an intermediary. We employ a simple game, which allows one player (the first mover) the option of undertaking an anti-social action (keeping money at the expense of the poorest player) and the option of whether to do this directly or through an intermediary. An unaffected party is told what happened and has the opportunity to punish the first mover. Though using the intermediary leads to much less equitable outcomes, and subjects expect this, they punish the first mover less when she elects to keep money via intermediation. This is true even if the intermediary is unambiguously free of responsibility. We then perform a treatment that encourages subjects to think and write about the game and the strategies of the other players before they play. 15 of 24 write down that the first mover will use the intermediary to avoid punishment. When these 15 subjects subsequently play the game, they punish intermediation less. Thus even when subjects understand the poor outcomes intermediation produces or the punishment-skirting intentions of the first mover, they punish in a manner suggesting they believe using the intermediary actually is less punishable, that something about this act is less outrageous. This result is particularly interesting for two main reasons: First, this cannot be explained by any current model or literature in Economics. Second, intermediation seems to be quite common and perhaps our moral intuitions and judgments in the case of individual or corporate partnerships should be questioned.

"Shallow are the souls that have forgotten how to shudder." - Kass (1997)[20]

The best explanations for the findings herein come from Moral Psychology. Moral judgment has been shown to be an emotional (Greene et al 2001), intuitive (Haidt 2001) response rather than a conscious, reasoned process; we judge something to be wrong because it just feels wrong and only try to justify with reasoning after the judgment has been made (Haidt et al 2000). Further, utilitarian judgments take longer to make than non-utilitarian judgments under cognitive load (Greene et al 2008b), and judgments become more utilitarian with subjects whose prefrontal cortex is damaged (Koenigs et al 2007), suggesting there are at least two parts of the brain operating in tandem in moral decision-making, one utilitarian and one driven by emotion and intuition, and that perhaps it is the emotional side that dominates this cerebral negotiation. Subsequently, it has been suggested and demonstrated that judgment and punishment is guided by outrage rather than the outcomes or intentions (Frederick & Kahneman 2002). Roth (2007) identifies a psychologically similar phenomenon in markets, which he labels "repugnance"; he finds that, "distaste for certain kinds of transactions can be a real constraint on markets." This distaste is not necessarily a function of the harm of the transaction but rather the repugnance felt; Perhaps it should not bother me if my neighbors were eating horse for dinner, but it might.

Outrage or repugnance seems to be attenuated in the cases of indirect or distant action, and such considerations are absent in current models. Royzman & Baron (2002) experimentally validate *the doctrine of the double effect* to be in line with subjects' moral preferences: Subjects judge an act to be more grievous when the harm is a means to an end rather than a side effect of the goal. In an interesting extension of Philippa Foot's trolley problem (1978), Greene et al (2008a) show that pushing a man in front of a moving trolley is more immoral if it is a direct result of the actor's physical force. This paper extends this understanding that bad acts are morally superior when proximity or directness decreases. Like Cox et al's (2008) inclusion of omission bias into their model of fairness, our models can be improved by including deontological considerations of fairness, not just utilitarian.

The results herein also disagree with the claim that norms are efficiency-enhancing. Norms can be efficiency-enhancing, but they need not be. The notion that norms are an evolutionary response to other tendencies such as self-interest or self-serving bias and serve to achieve a socially preferable outcome is not supported in the contexts in this paper. This result suggests the need for policy; we cannot assume society will marshal from within itself forces to make faltering actors return to

[20]Kass, Leon. 1997. "The Wisdom of Repugnance." *The New Republic*, June 2: 17-26.

behavior required for its proper functioning. The preference for indirect harm may lead to abated punishment in many arenas - jury sentencing, product boycotting, or shunning our friends. An increased awareness of our tendencies in these situations might help us understand and maybe correct the judgments we are otherwise inclined to make and lead to more efficient outcomes.

References

[1] Abreu, Dilip, David Pearce, and Ennio Stacchetti. 1986. "Optimal Cartel Equilibria with Imperfect Monitoring." *Journal of Economic Theory*, 39(1): 251-269.

[2] Axelrod, Robert. 1986. "An Evolutionary Approach to Norms." *American Political Science Review*, 80(4): 1095-1111.

[3] Bartling, Bjorn and Urs Fischbacher. 2008. "Shifting the Blame." SSRN Working Paper 1166544.

[4] Bénabou, Roland and Jean Tirole. 2006. "Incentives and Prosocial Behavior." *American Economic Review*, 96(5): 1652-1678.

[5] Bolton, Gary and Axel Ockenfels. 2000. "ERC: A Theory of Equity, Reciprocity and Competition." *American Economic Review*, 90(1): 166-193.

[6] Cain, Daylian; Loewenstein, George and Don A. Moore (2005) "The Dirt on Coming Clean," *Journal of Legal Studies*, 34(1), 1-25.

[7] Charness, Gary. 2000. "Responsibility and Effort in an Experimental Labor Market." *Journal of Economic Behavior & Organization*, 42: 375-384.

[8] Charness, Gary. 2004. "Attribution and Reciprocity in a Simulated Labor Market: An experimental investigation." *Journal of Labor Economics*, 22(3): 665-688.

[9] Charness, Gary and Martin Dufwenberg. 2006. "Promises and Partnerships." *Econometrica*, 74(6): 1579-1601.

[10] Charness, Gary and Matthew Rabin. 2002. "Understanding Social Preferences with Simple Tests." *Quarterly Journal of Economics*, 117: 817-869.

[11] Cox, James C., Daniel Friedman, and Vjollca Sadiraj. 2008. "Revealed Altruism." *Econometrica*, 76(1): 31-69.

[12] Cushman, Fiery. 2008. "Crime and Punishment: Distinguishing the roles of causal and intentional analyses in moral judgment." *Cognition*, 108: 353-380.

[13] Dufwenberg, Martin and Georg Kirchsteiger. 2004. "A Theory of Sequential Reciprocity." *Games and Economic Behavior*, 47: 268-298.

[14] Dyck, IJ Alexander and Luigi Zingales. 2002. "The Corporate Governance Role of the Media." NBER Working Paper no. W9309.

[15] Falk, Armin, Ernst Fehr, and Urs Fischbacher. 2005. "Driving Forces Behind Informal Sanctions." *Econometrica*, 73(6): 2017-2030.

[16] Falk, Armin; Ernst Fehr, and Urs Fischbacher. 2008. "Testing Theories of Fairness - Intentions Matter." *Games and Economic Behavior*, 62: 287-303.

[17] Falk, Armin and Urs Fischbacher. 2006. "A Theory of Reciprocity." *Games and Economic Behavior*, 54: 293-315.

[18] Fehr, Ernst and Urs Fischbacher. 2004. "Third Party Punishment and Social Norms." *Evolution and Human Behavior*, 25: 63-87.

[19] Fehr, Ernts and Simon Gächter. 2000. "Cooperation and Punishment in Public Goods Experiments." *American Economic Review*, 90: 980-994.

[20] Fehr, Ernst; Oliver Hart, and Christian Zehnder. 2008. "Contracts, Reference Points, and Competition." http://www.economics.harvard.edu/faculty/hart/files/Contracts, Reference Points and Competition.pdf

[21] Fehr, Ernst and Klaus M. Schmidt. 1999. "A Theory of Fariness, Competition, and Cooperation." *Quarterly Journal of Economics*, 114:817-868.

[22] Fershtman, Chaim and Uri Gneezy. 2001. "Strategic Delegation: An Experiment," *The RAND Journal of Economics*, 32(2): 352-368.

[23] Fischbacher, Urs. 1999. "z-Tree - Zurich Toolbox for Readymade Economic Experiments - Experimenter's Manual." Institute for Empirical Research in Economics, University of Zurich Working Paper Nr. 21.

[24] Fiske, Alan Page and Philip E. Tetlock. 1997. "Taboo Trade-offs: Reactions to Transactions That Transgress the Spheres of Justice." *Political Psychology*, 18(2): 255-297.

[25] Foot, Philippa. 1978. *Virtues and Vices and Other Essays*. Berkeley, California: University of California Press.

[26] Greene, Joshua D. and Jonathan Haidt. 2002. "How (and where) Does Moral Judgment Work?" *TRENDS in Cognitive Sciences*, 16(12): 517-523.

[27] Greene, Joshua, D. Lindsell, A.C. Clarke, L.E. Nystrom, J.D. and Cohen. 2008. "Pushing moral buttons: The interaction between personal force and intention in moral judgment" Unpublished Manuscript, Harvard University.

[28] Greene, Joshua, Sylvia Morelli, Sylvia, Kelly Lowenberg, Leigh Nystrom, and Jonathan Cohen. 2008. "Cognitive Load Selectively Interferes With Utilitarian Moral Judgment." *Cognition*, 107(3): 1144-1154.

[29] Greene, Joshua, Brian Sommerville, Leigh Nystrom, John Darly, and Jonathan Cohen. 2001. "An fMRI Investigation of Emotional Engagement in Moral Judgment." *Science*, 293: 2105-2108.

[30] Grossman, Zachary. 2009. "Self-Signaling Versus Social-Signaling in Giving." University of California Santa Barbara Economics Departmental Working Paper 01-09.

[31] Haidt, Jonathan. 2001. "The Moral Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment." *Psychological Review*, 108: 814-834.

[32] Haidt, Jonathan, F. Bjorklund, and S. Murphy. 2000. "Moral Dumbfounding: When Intuition Finds no Reason." Unpublished Manuscript, University of Virginia.

[33] Hamman, John, George Loewenstein, and Roberto A. Weber. 2007. "Self-interest Through Agency: An Alternative Rationale for the Principal-Agent Relationship." http://www.andrew.cmu.edu/user/rweber/Files/AgencyPaper2008-7-29.pdf

[34] Hirschman, Albert O. 1970. *Exit, Voice, and Loyalty*. Cambridge, MA: Harvard University Press.

[35] Kahneman, Daniel and Shane Frederick. 2002. "Representativeness Revisited: Attribute Substitution in Intuitive Judgment." In *Heuristics and Biases*; ed. Gilovich, Thomas, Dale Griffin, Daniel Kahneman, 49-81.

[36] Kahneman, Daniel, Jack Knetsch, and Richard Thaler. 1986. "Fairness as a Constraint on Profit-Seeking." *The American Economic Review*, 76(4): 728-741.

[37] Koenigs, Michael, Lian Young, Ralphs Adolph,Daniel Tranel, Fiery Cushman, Marc Hauser, and Antonio Damasio. 2007. "Damage to the Prefrontal Cortex Increases Utilitarian Moral Judgments." *Nature*, 446: 908-911.

[38] Martin, Kenneth and Randall Thomas. 2005. "When Is Enough, Enough? Market Reactions to Highly Dilutive Stock Option Plans and the Subsequent Impact on CEO Compensation." *Journal of Corporate Finance*, 11(1): 61-83.

[39] Paharia, Neeru, Karim Kassam, Joshua Greene, and Max Bazerman. 2007. "Washing Your Hands: Moral Implications of Indirect Action." http://www.people.hbs.edu/mbazerman/Papers/09-012.pdf

[40] Rabin, Matthew. 1993. "Incorporating Fairness into Game Thoery and Economics." *American Economic Review*, 83(5): 1281-1302.

[41] Roth, Alvin E. 2007. "Repugnance as a Constraint on Markets." *Journal of Economic Perspectives*, 21(3): 37-58.

[42] Royzman, E. and Jonathan Baron. 2002. "The Preference for Indirect Harm." *Social Justice Research*, 15: 165-184.

[43] Stigler, George J. 1964. "A Theory of Oligopoly." *Journal of Political Economy*, 72(1): 44-61.

[44] Sunstein, Cass. 2005. "Moral Heuristics." *Behavioral and Brain Sciences*. 28: 531-573.

Appendix

Table 6: Average Punishment for Each Scenario

Scenario	Sold?	Price	Amt. Kept	Avg. Punishment
1	No	–	\$5	0.69***
				(0.21)
2	No	–	\$6	1.30***
				(0.25)
3	No	–	\$7	2.03***
				(0.28)
4	No	–	\$8	2.89***
				(0.33)
5	No	–	\$9	4.00***
				(0.39)
6	No	–	\$10	5.22***
				(0.48)
7	Yes	\$5	\$5	0.78***
				(0.22)
8	Yes	\$5	\$6	0.89***
				(0.22)
9	Yes	\$5	\$7	1.06***
				(0.24)
10	Yes	\$5	\$8	1.03***
				(0.23)
11	Yes	\$5	\$9	1.03***
				(0.23)
12	Yes	\$5	\$10	1.27***
				(0.26)
13	Yes	\$6	\$6	1.19***
				(0.26)
14	Yes	\$6	\$7	1.36***
				(0.26)
15	Yes	\$6	\$8	1.45***
				(0.27)
16	Yes	\$6	\$9	1.59***
				(0.29)
17	Yes	\$6	\$10	1.44***

Continued on next page

Table 6 – continued from previous page

Scenario	Sold?	Price	Amt. Kept	Avg. Punishment
				(0.28)
18	Yes	$7	$7	2.02***
				(0.33)
19	Yes	$7	$8	2.05***
				(0.31)
20	Yes	$7	$9	1.94***
				(0.32)
21	Yes	$7	$10	1.91***
				(0.32)
22	Yes	$8	$8	2.50***
				(0.35)
23	Yes	$8	$9	2.38***
				(0.36)
24	Yes	$8	$10	2.56***
				(0.36)
25	Yes	$9	$9	3.59***
				(0.43)
26	Yes	$9	$10	3.45***
				(0.42)
27	Yes	$10	$10	4.30***
				(0.49)

Table 7: Summary Statistics

	Intermediation Game	**No Intermediary Treatment**
No. Subjects	64	24
No. Who Use Intermediary	33	-
Avg. Sell Price	\$7.17	-
Avg. Amount Kept in \$10 DG	\$8.37	\$5.50
(when sold)	\$9.79	-
(when unsold)	\$6.87	-
% Who Punish At Least Once	81%	100%
Avg. Punishment	\$2.07	\$3.65
Avg. Nonzero Punishment	\$4.24	\$4.70
Avg. Punishment as % of first mover's pre-punishment wealth	28%	44%
Avg. Nonzero Punishment as % of first mover's pre-punishment wealth	58%	56%
Avg. Proft (All Subjects)	\$4.31	\$4.83

Table 8: What Punishment Amounts Were Common?

	All	Direct	Indirect
No. Reported Punishments	1,728	384	1,344
Punishment = \$0	884	157	727
Punish Everything (to \$0)	253	57	196
Equate w/ Punisher (to \$5)	188	64	124
Equate w/ Receiver	129	44	85
Equate w/ Intermediary	92	NA	92

When a punishment qualifies for more than one row, it is only included in the first row for which it qualifies.

Table 9: How Many Subjects Punish Directness More Harshly[†]

No. Subjects Who...	
Punish Direct Action Harsher More Frequently	28
Punish Direct Action Harsher At Least Once	36
Never Punish Indirect Action Harsher	46
Total No. Subjects	64

Table 10: Expected Payouts

Pre-punishment Wealth	Exp. Payout Unsold	Exp. Payout Sold	Ranksum p-value[a]	Permutation p-value[a]
\$10	\$4.78	\$5.70	<0.01	0.01
\$9	\$5.00	\$5.41	0.05	0.07
\$8	\$5.11	\$5.44	0.07	0.06
\$7	\$4.97	\$4.96	0.61	1.00
\$6	\$4.70	\$4.41	0.48	0.13
\$5	\$4.31	\$3.73	<0.01	<0.01

Expected Payouts in "sold" column are "worst case" scenarios -
They assume the intermediary will take the action that maximizes punishment.
[a]Both tests are two-tailed. Permutation p based on 200,000 permutations.

6.1 Order Effects and Within-Subject Consistency

The subjects played the game for four periods, in all four roles. This begs two questions. First, does the order of roles played affected decisions as a punisher? Second, do subjects show behavioral correlations across roles? We find no evidence of order effects with respect to the main result, but we do find within-subject correlations.

Role order does not affect how subjects punished indirectness. Table 12 includes three dummy variables (Played as A, B or C) which equal 1 if at the time the subject was a punisher, she had already played the game in the role specified[21]. This specification shows that how much punishment decreases when the game is sold is not affected by these order effects. The coefficients of the three-way interactions in the table identify how much punishment decreases when the DG is sold for punishers who played the game in each of the other roles first. None of these coefficients are statistically different than zero indicating that the punishment decrease is not impacted by the

[21]Subjects all played the game in the order $D \rightarrow C \rightarrow B \rightarrow A \rightarrow D$. Thus we do not have any data for subjects who only played as C or B before they punished; they always played as A as well.

Table 11: OLS Regressions

Subjects Who Expect Intermediation to Hurt Equity N=46		
Dep. Var. = Punishment ($)		
	All Scenarios (I)	Intermediary Makes $0 (II)
First Mover's Profit ($) over $5	0.91***	0.91***
	(0.11)	(0.11)
First Mover's Profit over $5 *Sold	-0.24***	-0.14**
	(0.05)	(0.05)
Intermediary's Profit	0.03**	–
	(0.02)	
Sold Dummy	0.26***	0.08
	(0.09)	(0.09)
Period	-0.06	-0.05
	(0.30)	(0.34)
Constant	1.02	1.02
	(1.15)	(1.29)
Session F.E.	Included	Included
Standard Errors Clustered at Subject Level		
***, **, * indicates p<0.01, 0.05, 0.1 respectively		

order of roles the subject experienced. Another way to read the table, is that in Period 1, when the punisher is untainted by experience, there is still a significant decrease in punishment for money made indirectly. This is evidenced in the statistically significant coefficient of the interaction of first mover's profit and whether she sold the DG.

Table 13 shows evidence of some behavioral consistency across roles. This table runs OLS regressions interacting the two main drivers of punishment - 'Player A's profit above $5' and 'Player A's profit when the game was sold' - with (i) whether the punisher sold the DG when she was player A and separately with (ii) how much profit above $5 she claimed for herself as Player A. Model I includes the data from all players. The first three rows show, respectively, punishment increases as Player A's profit increases, this is not different for punishers who sold the DG, but it is less for punishers who claimed (or will claim) more money when they were Player A. That is, they do not punishing selfishness as harshly if they were (or are going to be) selfish themselves. The next three rows, respectively, show that punishment of profit made decreases if made by selling (Result 1 in the paper), this decrease is much greater for punishers who sold the game themselves, and it is no different for punishers who made (or will make) a lot of money as Player A. In other words, subjects who take (or will take) the indirect action, punish indirectness less harshly.

Table 12: OLS Regressions: Order Effects

Dep. Var. = Punishment ($)	
First Mover's Profit ($) over $5	0.90*** (0.09)
First Mover's Profit over $5 *Sold	-0.21*** (0.03)
First Mover's Profit over $5 *Sold Played as A first	-0.07 (0.16)
First Mover's Profit over $5 *Sold Played as B first	0.01 (0.17)
First Mover's Profit over $5 *Sold Played as C first	0.00 (0.14)
Intermediary's Profit	0.05** (0.02)
Sold Dummy	0.31*** (0.08)
Played as A first	-0.77 (0.72)
Played as B first	-0.04 (0.57)
Played as C first	-0.02 (0.56)
Constant	1.73** (0.77)
Session F.E.	Included

Standard Errors Clustered at Subject Level
***, **, * indicates p<0.01, 0.05, 0.1 respectively

6.2 Do Subjects Not Account for the Intermediary's Endowment?

There is reasonable concern that the punishers do not account for the intermediary's $5 endowment when making their fairness judgments. If the punisher thinks both the intermediary and the receiver have $0, then "selling" the DG can be construed as a pro-social move, granting a poor player a chance to receive some wealth. This would drive similar punishment patterns as we observe. Thus let us consider only scenarios where the first mover does not leave enough for money to be shared amongst the other two players. If the first mover chooses to make $9 or $10, then there is $1 or $0 left over. Since dollars are indivisible in this game, selling the game for $9 or keeping $9 in the DG should be identical: The first mover has chosen a final allocation of ($9, $1, $0) or ($9, $0, $1). To keep the outcome constant, let us only consider scenarios where the intermediary passes the dollar to the receiver, so the final allocation is ($9, $0, $1). We shall also restrict the analysis to punishers who do not (mistakenly) shift any blame to the intermediary in any of these scenarios. That is, we only include punishers who do not punish the intermediary in the four scenarios we have isolated

Table 13: OLS Regressions: Behavioral Consistency

Dep. Var. = Punishment ($)			
	All	Sold as A	Didn't Sell as A
	(I)	(II)	(III)
First Mover's Profit ($) over $5	1.62***	1.71**	1.64***
	(0.38)	(0.67)	(0.49)
First Mover's Profit over $5 *Sold DG as Player A	0.05	–	–
	(0.18)		
First Mover's Profit over $5 *Profit made as A	-0.10*	-0.11	-0.09
	(0.05)	(0.08)	(0.07)
First Mover's Profit over $5 *Sold	-0.39*	-0.57	-0.42*
	(0.23)	(0.53)	(0.24)
First Mover's Profit over $5 *Sold *Sold DG as Player A	-0.16*	–	–
	(0.09)		
First Mover's Profit over $5 *Sold *Profit made as A	0.03	0.04	0.02
	(0.03)	(0.07)	(0.03)
Intermediary's Profit	0.05**	0.04**	0.07
	(0.02)	(0.02)	(0.05)
Sold Dummy	0.31***	0.15***	0.48***
	(0.08)	(0.06)	(0.14)
Sold DG as Player A	-1.06	–	–
	(0.49)		
Profit made as A	0.37**	0.41*	0.29
	(0.15)	(0.23)	(0.22)
Period	-0.20	-0.06	-0.30
	(0.25)	(0.31)	(0.38)
Constant	-0.88	0.13	-3.07
	(1.46)	(3.31)	(2.88)
Session F.E.	Included	Included	Included
Observations	64	33	31
Standard Errors Clustered at Subject Level			
***, **, * indicates p<0.01, 0.05, 0.1 respectively			

- Sell for $10, Don't Sell and Keep $10, Sell for $9 and intermediary passes the $1, and Don't Sell and Keep $9.

Restricting our analysis thusly does not change the punishment comparisons. A two-tailed, signed rank test rejects the hypothesis that the punishments are the same whether the first mover sells or not with a p=0.02. Similarly, a Fisher Pitman matched pair permutation test rejects the same hypothesis at the same level (p=0.02).

6.3 Study 2 Appendix

Table 14: Survey 1 Summary Stats & Randomization

	Pooled	Direct	Indirect
Observations	202	108	94
Gender (Male=1, Female=0)	0.31	0.33	0.35
Freshman	0.33	0.36	0.30
Sophomore	0.32	0.33	0.30
Junior	0.31	0.27	0.36
Senior	0.03	0.04	0.02
Member of Charity Group (PBHA)	0.59	0.60	0.59

Table 15: Survey 2 Summary Stats & Randomization

	Pooled	Direct, Reward	Indirect, Reward	Direct, Guess	Indirect, Guess
Observations	135	42	31	31	31
Gender (Male=1)	0.35	0.40	0.29	0.35	0.35
Freshman	0.30	0.21	0.39	0.26	0.39
Sophomore	0.27	0.36	0.29	0.23	0.16
Junior	0.39	0.31	0.32	0.52	0.42
Senior	0.02	0.05	0.00	0.00	0.03
PBHA	0.64	0.67	0.65	0.68	0.55

Experiment Instructions

Preliminaries

Thank you for participating in this study. If at any time you have questions, please raise your hand, and I will assist you. From now until the end of the session, communication of any kind between participants is not allowed. Please do not use the computer for any other purpose than participating in this study. Also, please turn off your cell phones.

Overview

The purpose of this session is to study decision making.
You will make a series of decisions that will affect your payoffs as well as the payoffs of others.
All studies run here in the Computer Lab for Experimental Research (CLER) do NOT use deception. Everything in these instructions is true, including the rules and the payoffs. Everyone received the same instructions.

What are my payoffs?

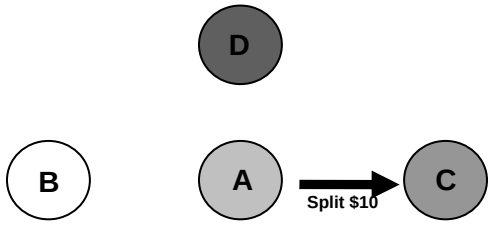
You will be paid your $10 show up fee plus the money made during the course of the following game.
As an example just for illustrative purposes: If you make $5 in the game, you will be paid $15 total.

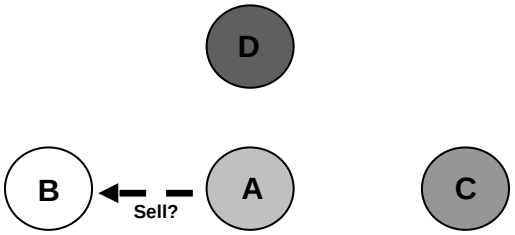
The Game

The game will be played in groups of four. We will play the game four times. Each time we play the game will be referred to as a "period".
In each period, everyone will be randomly assigned to one role and one group. You will play each role exactly once.
There are no computer players. The other three players in your group will be three other people in the room. You will be playing with three different people in each period.

Let's call the four roles A, B, C, and D. The game is completely anonymous, so players will only be referred to as A, B, C, or D for the duration of the experiment.

Here is how the game works:
Player A owns the rights to split $10 with C. That is, A decides how to split the $10, and the two players are paid accordingly. C has no say in the decision. C will simply be informed of A's anonymous decision.

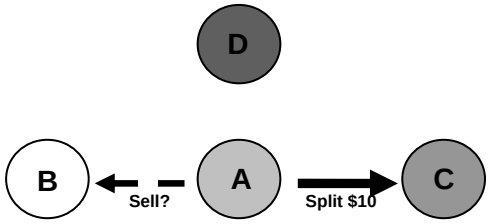


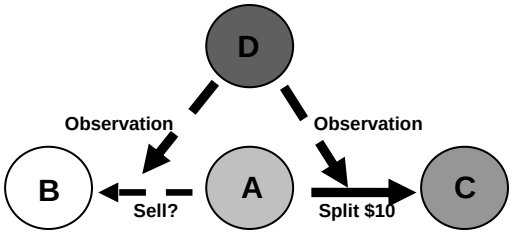


Before A splits the $10 however, A can "sell" the rights to play this game to player B.
A chooses whether to sell, and how much to sell for. B has no say in the decision.

If the game is "sold" to player B, then B splits $10 with C.
If the game is not "sold", then A splits $10 with C.
B's profit is whatever she keeps on top of the price she paid for the "Split $10" Game.

Also, Player B automatically has $5 in addition to any profit she earns in the game.

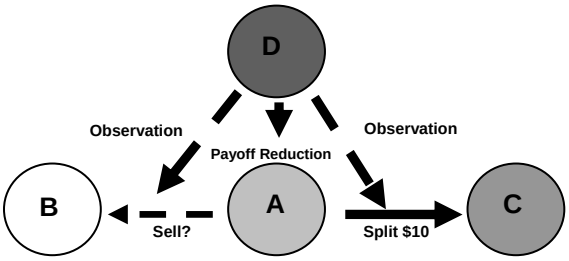




Player D then has the option of reducing Player A's payoff. They get to make this decision based on what happened in the game:
1. Whether the game was sold
2. If so, for how much
3. How was the $10 split with Player C

Player D can reduce Player A's payoff by any amount; However, she cannot reduce Player A to negative amounts.
So say Player A makes $7 (either by selling or playing the "Split $10" Game). Player D can punish any amount between $0 and $7.

Player D earns exactly $5 for the period no matter what punishment decisions she makes.



At the end of each period, you will not be told the decisions of the other players. We will learn what happened once we have played all four periods.
At that point, the computer will randomly choose one period. Your profits from that period, and that period only, will be the amount that you are paid when you leave, in addition to your show up fee.
Everyone will be paid for the same period. All four periods are equally likely to be the "payment period", so make thoughtful decisions in all four periods.

To recap, broken down by roles, the game is as follows:

Player A

- Owns the rights to anonymously split $10 with C, without any input from C
- Can "sell" this to B
- Can be punished by D
- If A sells the Split $10 game to B, Profit is equal to the price for which the game sells minus the punishment decided by D.
- If A does not sell the Split $10 game, Profit is equal to the amount A keeps in the "Split $10" game minus the punishment decided by D.

Player B

- Can purchase from A the rights to anonymously split $10 with C, without any input from C
- Cannot be punished by D
- If B is sold the Split $10 game by A, Profit is equal to the amount kept in the game minus the price paid for the game.
- If B is not sold the Split $10 game by A, Profit is 0.
- Will be paid $5 plus any profits made in this game.

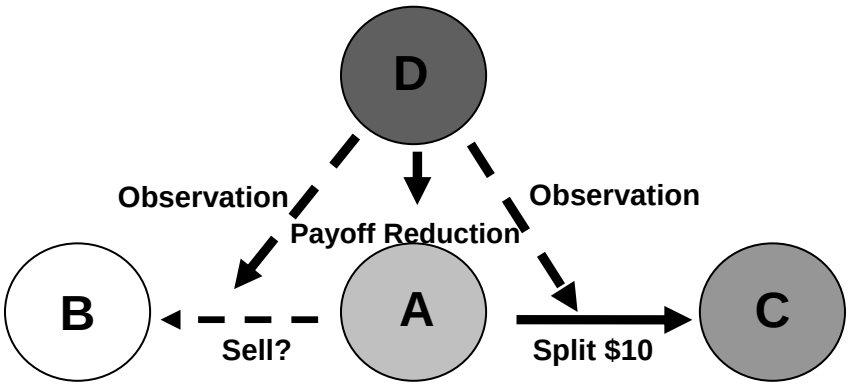
Player C

- Receives the amount that is decided by A or B. This is C's Profit.

Player D

- Based on the behavior of A and B, D can reduce A's payoff by any amount all the way to $0.
- Profit equals $5 no matter what.

Are there any questions about how this game works?



If at any time during the session you have questions, please raise your hand.

Good luck.