

BARGAINING STRUCTURE, FAIRNESS, AND EFFICIENCY

Matthew Rabin
Department of Economics
University of California, Berkeley

First Draft: June 15, 1996
This Draft: February 24, 1997

Abstract: Experiments with the ultimatum game—where one party can make a take-it-or-leave-it offer to a second party on how to split a pie—illustrate that conventional game theory has been wrong in its predictions regarding the simplest of bargaining settings: Even when one party has enormous bargaining power, she may be able to extract all the surplus from trade, because the second party will reject grossly unequal proposals. But ultimatum games may lead us to misconstrue some general lessons: Given plausible assumptions about what preferences underlie ultimatum-game behavior, alternative bargaining structures that also give a Proposer enormous bargaining power may lead to very different outcomes. For virtually any outcome in which the Proposer gets more than half the pie, there exists a bargaining structure yielding that outcome. Notably, many bargaining structures can lead to inefficiency even under complete information. Moreover, inefficiency is partly caused by asymmetric bargaining power, so that “fairer environments” can lead to more efficient outcomes. Results characterize how other features of simple bargaining structures affect the efficiency and distribution of bargaining outcomes, and generate testable hypotheses for simple non-ultimatum bargaining games.

Keywords: Bargaining, Efficiency, Fairness, Inefficiency, Inequality, Ultimatum Game

JEL Classification: A12, A13, B49, C70, D63

Acknowledgments: I thank seminar participants at Berkeley, University of Arizona, Johns Hopkins, University of Maryland, MacArthur Foundation Costs-of-Inequality meetings, Gary Charness, Jeff Dayton-Johnson, and Erik Eyster for helpful comments. I am also grateful to Erik Eyster and Clara Wang for research assistance, to Paul Smolensky for a crucial cappuccino, and the Russell Sage, Alfred P. Sloan, and MacArthur Foundations for financial support related to this project. This draft still sucks, so comments would be appreciated. They can be directed to: Department of Economics/549 Evans Hall #3880/University of California—Berkeley/Berkeley, CA 94720-3880. E-mail: rabin@econ.berkeley.edu. CB handle: “Game Boy.”

1. Introduction

Everyday knowledge tells us that humans are inclined to retaliate against those who treat them unfairly. This fact has been verified repeatedly in laboratory experiments, most commonly by having subjects play variants of “the ultimatum game.” In ultimatum games, the Proposer makes a take-it-or-leave-it offer to the Decider on how they should split some money. The Decider then chooses whether to accept or reject that split. If she accepts, the two get the proposed allocation. If she rejects, they each get nothing. Classical game theory—presuming self-interested behavior as narrowly defined—predicts that the Proposer will offer (virtually) nothing to the Decider, and that the Decider will accept any positive offer. But Proposers rarely propose to keep nearly all of the pie for themselves, and extremely lopsided offers are typically rejected when made. Proposers often propose 50/50 splits; for pie sizes which have been investigated (typically between \$5 and \$20, and as high as \$100), offers to the Decider average about 40% of the pie.¹

Results from the ultimatum game show that bargaining power will not be fully exploited, and that division of surplus will be more even than in conventional game theory assuming self-interested players. But when there is complete information, there is every reason to expect that the actual outcome will accord with the conventional game-theoretic prediction in one respect: It will be efficient. If the Proposer is fully aware of what the Decider’s “threshold” is, she will always make an offer sharing the entire surplus that the Decider will accept.

In this paper, I argue that the ultimatum game is somewhat special in its prediction of efficiency. With a broad range of plausible preferences that explain ultimatum-game behavior, virtually any outcome where the Proposer gets more than half the pie—including inefficient outcomes—is the unique equilibrium outcome of some perfect- and complete-information bargaining game. All the bargaining games considered would predict full efficiency for fully self-interested preferences. Hence, obscured by the special features of the ultimatum game is the fact that realistic departures from self interest may lead to inefficiencies in bargaining. I show that inefficiencies are especially likely in situations where bargaining power is very lopsided in favor of the Proposer, and where the Decider can cheaply impose a moderate amount of harm on the Proposer.

The logic for how inefficiencies arise is simple. A person who is treated unfairly in bargaining will retaliate against the unfair party. Hence, unfair offers in bargaining will tend to generate responses leading to inefficient outcomes. In some settings, such as the ultimatum game, the threatened inefficiency will sufficiently deter unfair offers to prevent retaliation, and an efficient outcome will ensue. But in other settings, a person with bargaining power will choose to accept retaliation rather than make an offer sufficiently generous to deter retaliation. The price for non-retaliation may be more than powerful people are willing to pay. The logic for how inefficiency arises makes clear that it results from unequal bargaining power: If a party were unable to demand much more than half the pie, then he would prefer to make an acceptable fair offer rather than suffer any substantial retaliation.

Consider the following simplified version of the ultimatum game:

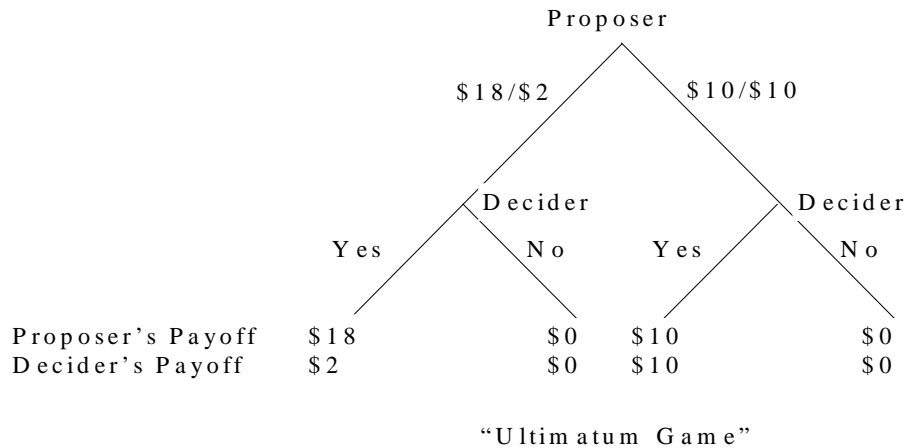


Figure 1

The Proposer can offer to split \$20 either by giving the Decider \$10, or by giving the Decider only \$2. (In actual ultimatum games, the Proposer has available a broader range of offers.) What will happen? Classical game theory, assuming common knowledge of rational self interest, predicts that the Proposer will offer the \$18/\$2 split, and the Decider will accept this. In reality, however, the majority of the Deciders would reject the \$18/\$2 offer. Their true payoffs do not simply reflect their monetary gains: An \$18/\$2 offer is intentionally and egregiously unfair, and many Deciders would prefer to forego \$2 rather than let Proposers get away with such a division. Since most Proposers would correctly anticipate this behavior, and also know that Deciders would accept the \$10/\$10 offer, even potentially unfair Proposers will perceive it as in their self-interest to offer \$10/\$10 rather than \$18/\$2. Occurrences of rejected \$18/\$2 offers would almost surely be due to incorrect expectations by Proposers who don't predict that the Decider they face will reject such offers. The only occurrences of inefficiency in ultimatum games, therefore, are because of incomplete information.

Now consider Figure 2, the "Sabotage Game". The Proposer is again in a position to offer either a \$18/\$2 split or a \$10/\$10 split. However, now the Decider does not have the option of rejecting either offer outright. Indeed, she chooses whether to engage in costly "sabotage" against the Proposer. This game might capture a situation where an employee can't or won't turn down an unfair offer from her more powerful employer, but where she'll have a later opportunity to engage in low-cost retaliation.

¹ See Guth, Schmittberger and Schwarze (1982), Guth and Tietz (1990) Thaler (1988), and Camerer and Thaler (1995) for analysis of the ultimatum game. See Roth (1995) for a review of experimental bargaining games more generally.

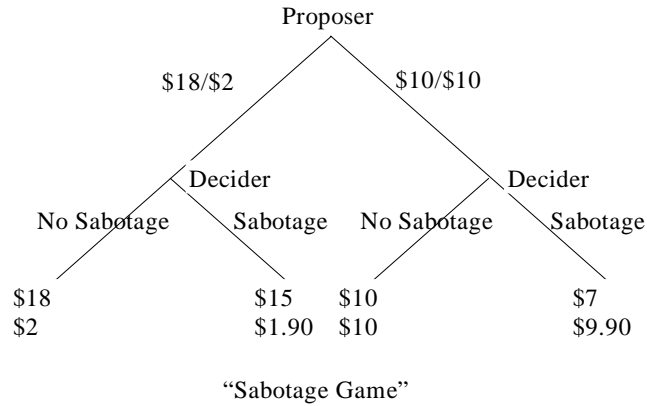


Figure 2

What is likely to happen in this game? As with the ultimatum game, a \$10/\$10 offer will be accepted—it is manifestly fair. This leaves the questions of how the Decider will react to an offer of \$18/\$2, and what the Proposer will do given his beliefs about her reaction. We can conjecture that many of the Deciders will “reject” the offer for much the same reason they do so in ultimatum games. Now, at a cost of 10 cents, the Decider can punish the Proposer by \$3 for making an unfair offer.

If the Proposer predicts that \$18/\$2 will be “rejected”, what will he do? Proposers innately motivated by fairness will propose \$10/\$10 here, just as in the ultimatum game. But those out to maximize their money payoffs are clearly going to offer \$18/\$2 here: No matter how the Decider responds, offering \$18/\$2 is a better choice for the Proposer. The Proposer can anticipate that the Decider will retaliate against him for his unjust proposal, and yet still he will make the unjust proposal. The cost of avoiding retaliation—making a fair offer—is too high.

Because many Proposers may be motivated by fairness, it is unclear what most Proposers will do. But especially for huge stakes, it is likely that many Proposers would be tempted to make the ungenerous offer, and live with the retaliation. As such, an inefficient outcome is likely. The inefficiency here is perhaps surprising from a traditional economic perspective, because all the usual suspects for inefficiency are absent: This is a complete- and perfect-information game, with no “transactions costs” of a classical sort.

This paper formalizes a simple class of bargaining games and the type of preferences that seemingly underlie the behavior posited in Figures 1 and 2 in order to explore some possible general resulting from departures from self interest in bargaining situations. Figure 3, which contains a generalization of Figures 1 and 2, helps illustrate some of the main results of this paper.

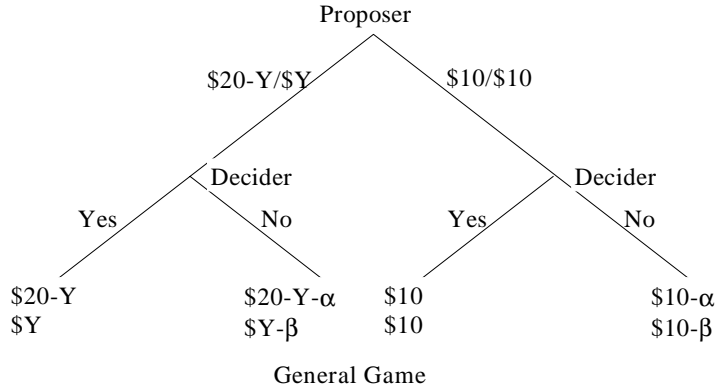


Figure 3

In Figure 3, assume that $Y < \$10$, α , and β are all positive. For ease of reference—this point is not crucial—assume that the parameters are such that both parties’ money payoffs for each outcome are non-negative, so that the true money payoffs are the maximum of zero and the values drawn. Given this, the ultimatum game corresponds to the case where $Y = 2$ and $\beta = \alpha = 18$. The sabotage game corresponds to the case where $Y = 2$ and $\alpha = 3$ and $\beta = .1$.

What outcomes are possible in such general games? The “payoffs” drawn in Figures 1, 2, and 3 are simply the money outcomes, whereas people’s true preferences include a taste for retaliation against unfair behavior. Most of the paper’s results concern predictions that hold for a very broad class of such retaliatory preferences. Proposition 1 below states the result that for *any* monetary outcome in which the Proposer gets at least \$10, there exists some combination of Y , α , and β such that this outcome is the unique equilibrium outcome when the Proposer is self-interested and the Decider has any taste for retaliation whatsoever. Choose any money payoff for the Proposer $z_1 > \$10$, and any payoff for the Decider z_2 such that $z_1 + z_2 \leq \$20$. Then choose $Y = z_2 + \epsilon$ and $\beta = \epsilon$, where ϵ is very small. Since $Y < \$10$, the Decider will have some taste for retaliation; since the cost of retaliation, ϵ , is very small, she will follow through on this taste. Then choose α so that $z_1 = 20 - Y - \alpha$. The point is that any outcome in which the Proposer gets more than half the pie can occur by the Proposer making an unfair offer, and the Decider sacrificing an arbitrarily small amount to punish the Proposer. The Proposer, if predominately self-interested, will live with the retaliation rather than make the fair \$10/\$10 offer.

Hence, for any outcome (z_1, z_2) , $z_1 > \$10$, we have found a bargaining structure generating that outcome. Therefore, *no matter* the precise nature of the Decider’s tastes, without knowing the exact bargaining structure we know virtually nothing about the outcome of the game.

The paper’s remaining results essentially explore what values of Y , α , and β tend to lead to an efficient outcome, where no surplus is dissipated. If either α or β is positive, an outcome in this class of games is efficient if and only if it involves the Decider saying “Yes” to the offer the Proposer makes in equilibrium. For what values of Y , α , and β will the outcome be efficient?

Since $(\$20 - Y - \alpha, \$Y - \beta)$ is the worst plausible outcome that can happen in this game, we know that if $Y + \alpha$ is very small, then the outcome will be relatively efficient—the Decider’s payoff may be driven down to zero, but the

Proposer will do quite well. Essentially, $20 - Y - \alpha$ represents the most favorable outcome that the Proposer can *force* on the Decider without the Decider's consent. The first simple efficiency result, therefore, is that whenever the Proposer cannot be severely punished by the Decider, the outcome will be relatively efficient.

The second efficiency result concerns the case Y is close to $\$10$, meaning the Proposer is simply unable to make a very unfair offer; the Proposer does not have much "bargaining power". In fact, if Y is close to $\$10$ the Proposer will only make the offer $(20 - Y, Y)$ if either α is very small, or if he is not worried that the Decider will actually punish him. So inefficiency can only arise if α is small *and* the Decider is willing to punish the Proposer. But as long as there is *some* upper bound on how much the Decider is willing to sacrifice to punish the Proposer by only a small amount, then almost surely the Decider will not punish the Proposer. Consequently, whenever Y is close to $\$10$, the equilibrium outcome will be relatively efficient, irrespective of the values of α and β , because retaliation will only be an equilibrium outcome if both values are quite small. In this sense, the outcome will be efficient whenever the Proposer does not have much bargaining power. Extrapolating from the simple class of games analyzed in this paper, we can say that inefficiencies are caused in part by asymmetric bargaining power: If a party does not have much bargaining power, he will not try to grab all of the pie, and hence revenge won't occur on the equilibrium path.

Neither of these results explains why the outcome is likely to be efficient in the ultimatum game with complete information, because in the ultimatum game of Figure 1, Y is small and α and β are large. Why is the outcome efficient? The answer is that the *sole* punishment available to the Decider is *so* severe to the Proposer: $\alpha = 20$. That is, if the Proposer is punished at all in the ultimatum game, he is punished so severely that he will surely be deterred from making an offer he knows will be rejected. This is in contrast to the sabotage game, where the punishment for an unfair offer is merely $\$3$, which may be tolerable for the Proposer. I define this severe-or-nothing feature of the punishment opportunities as "starkness", and formalize a simple result that the starkness of a bargaining game helps determine the efficiency of its outcome.

While this paper neither presents experimental results nor proposes specific experimental designs, it generates testable hypotheses for experimentalists, and illustrates features of non-ultimatum games that can help organize experimental research on bargaining situations. The formal model of this paper covers a simple, restrictive, and artificial class of bargaining games. The well-specified set of simple games has an obvious advantage for experimental tests. By moving beyond the ultimatum game, we can move beyond its potentially misleading lessons; by limiting ourselves to simple, two-move structures, however, we can avoid the difficulties in interpreting behavior in complicated, sequential-bargaining structures.

The intuition from the analysis in this paper may also help explain inefficiencies in a variety of real-world settings. In labor relations, a corporation may know that ramming through an unfair labor contract will lead to a strike, consumer boycott, bad P.R., or workaday retaliation by its workers, all leading to inefficiencies. But paying fair wages may be so costly that the corporation may just accept the inefficiency rather than forego the lion's share of profits. This observation is related to a result, often unemphasized, from the efficiency-wage literature: The wages set in a competitive labor market may be inefficiently low. Here, I am arguing that if workers have weak bargaining

power, this may lead to lower productivity than if they have strong bargaining power. This seems a natural corollary to the fairness-based efficiency-wage literature of the sort explored in Akerlof (1984) and Akerlof and Yellen (1990).²

There is a famous argument that the allocation of bargaining power induced by the legal regime will not, absent transactions costs, affect efficiency. The models of this paper suggest that this “Coase Theorem” result often may not hold in any useful sense. It is commonly argued that when there is complete information and court is costly, parties will find a mutually advantageous way to avoid court, no matter what the court will rule. This isn’t correct, however, if one of the litigants *likes* the fact that court is costly for the other party. She may be willing to pay court costs herself even if she is aware that she will not prevail in court. By contrast, if it is common knowledge that the courts will uphold the “fair” allocation, then the intuition that parties will settle disputes before going to court will again hold. This yields a simple prediction: Fair laws will be, *ceteris paribus*, more efficient than unfair ones.³ These issues apply to both civil laws, and labor laws—e.g., it may enhance efficiency for government to pass laws enhancing the bargaining power of labor unions, as this may decrease strikes and inefficient work rules.

In the next section, I define the class of bargaining structures I consider in this paper; in Section 3, I formalize the preferences, and in Section 4 I present general results. I then present an extended efficiency-wage example in Section 5, and conclude in Section 6.

2. Proposer-Decider Games

Consider a simple class of two-person bargaining games: One party (the Proposer) makes an offer on how to split some surplus, and a second party (the Decider) determines by her response whether and how the surplus is divided. The ultimatum game is an example of such a game: With a pie of size 1, the Proposer proposes some split of $(1-x, x)$, $x \in [0, 1]$. The Decider can respond “Yes” or “No.” If the Decider says “Yes,” then the game ends with the Proposer getting $\$(1-x)$ and the Decider getting $\$x$. If the Decider says “No,” they each get $\$0$. The dictator game is an even simpler example of a Proposer-Decider game: The Decider’s response doesn’t matter—whatever $(1-x, x)$ the Proposer proposes, that is the split.

I define a class of *Proposer-Decider Games* that has this same offer-response structure:

Definition 1: A game splitting a pie of size 1 is a *Proposer-Decider Game* (PDG) if it involves the following procedure:

The Proposer can propose any $(1-x, x)$ such that $x \in [Y, 1]$, where $Y \in [0, 1]$ is an exogenous parameter;

² Indeed, there is one reason to expect that the issues may be magnified in business/labor bargaining: Often one person’s decision whether to retaliate against a firm is mediated by those around him. You can either get away with shirking and sabotaging on the job or not—depending on how your coworkers feel about it. A multiplier effect arises—if you mistreat your employees, you not only have the direct effect on each employee, but the overall loss of social enforcement of civil behavior. There is sociological evidence (e.g. Mars (1982)) that group solidarity can be a serious problem for an organization that has lost the loyalty of its members, because it causes people to collude in retaliation.

³ The requirement that losing parties pay court costs can also be seen to be important here: Corresponding to the result that if $Y+\alpha$ is small we will get full efficiency, if the law fully compensates winning parties for all costs of going to court, then mistreated parties *can’t* use lawsuits as a punishment device. Hence, to be efficient, unfair laws must be backed up with losers-pay-court-costs provisions.

After observing x , the Decider chooses some $\gamma \in [0,1]$;

The Proposer and the Decider get monetary outcomes $g_1(1-x,\gamma) \geq 0$ and $g_2(x,\gamma) \geq 0$, respectively, where $g_1(\varrho,\varrho)$ and $g_2(\varrho,\varrho)$ are functions such that for all x : $g_1(x,0) = 1-x$ and $g_2(x,0) = x$, g_1 and g_2 are non-increasing in γ for all γ , and for all γ and γ' such that $g_1(x,\gamma) < g_1(x,\gamma')$, $g_2(x,\gamma) < g_2(x,\gamma')$.

In Proposer-Decider games, the Proposer proposes a split, where the rules of the game specify that he is not allowed to offer the Decider an amount less than Y . The Decider can choose either to accept the Proposer's offer or (according to some exogenously given rules) to "shrink" the Proposer's offer by some amount. By shrinking, I mean that the Decider can only choose to lower the Proposer's payoff, but only by lowering her own. The choice variable γ represents how much the Decider shrinks the pie— $\gamma = 0$ means not at all, $\gamma = 1$ means the maximum. So in addition to the take-it-or-leave-it nature of the ultimatum game, or the take-it nature of the dictator game, we can allow an infinite variety of take-some-of-it, leave-some-of-it games. The rules for shrinking are determined completely by the functions g_1 and g_2 .⁴

I study these games because they are the simplest class of games—two players, each moving once in sequence—that allow some richness to the bargaining structure. I now consider ways of measuring features of these games that will be of special importance. The value Y can be seen to determine to a large degree the Decider's "bargaining power." To clarify the role of Y and facilitate discussion, I separately define it:

Definition 2: A PDG has *decider power* ρ if $\rho \equiv \sup_{k \in [0,1]} \{k \mid \text{for all offers } x \geq Y, \text{ there exists a } \gamma \text{ such that } g_2(x,\gamma) \geq k\}$.

Decider power says that the Decider can always guarantee herself at least proportion ρ of the pie. Trivially, $\rho = Y$ in all PDGs. In any PDG, the conventional prediction assuming pure self-interest is for the Proposer to propose $(1-Y,Y)$, and the Decider to accept such an offer:

Lemma 1: If both the Proposer and the Decider were purely self-interested, the unique subgame-perfect Nash Equilibrium outcome would always be that the Proposer proposes $(1-Y,Y)$, and the Decider accepts this offer (chooses γ such that $(g_1(Y,\gamma),g_2(Y,\gamma))=(1-Y,Y)$).

Proof: Trivial. Q.E.D.

Lemma 1 provides a sense in which ρ can be seen as a measure of bargaining power. Of special interest is that $\rho = 0$ always implies that the unique subgame-perfect equilibrium when each player cares only about monetary gain is for the Decider to get zero. For both the ultimatum game and the dictator game, $\rho = 0$.

While ρ completely determines the outcome for pure self interest, it does not completely determine the outcome for the alternative preferences discussed in this paper. A second feature of interest is the degree to which the Proposer can "force" a particular outcome. It may be that the Decider has no say at all. This is the case in the

⁴ Fixing a monetary payoff for the Proposer, for the preferences defined in the next section the Decider would always choose the maximum monetary payoff for herself. So, it is without loss of generality that g_1 and g_2 are functions generating a one-dimensional locus of points, rather than being correspondences that allow a range of g_1 for each g_2 .

dictator “game”, where for all x and γ , $g_1(x, \gamma) = x$ and $g_2(x, \gamma) = 1 - x$, so that the Proposer can force an outcome of $(1,0)$. More generally, it is of interest what type of outcomes a Proposer is able to force without the Decider's cooperation:

Definition 3: A PDG has *force* π if $\pi \equiv \sup_{k \in [0,1]} \{k \mid \text{there exists } x \text{ such that } g_1(x, \gamma) \geq k \text{ for all } \gamma\}$.

To have a force of π means that the Proposer has a strategy that guarantees himself, *irrespective of the Decider's response*, proportion π of the pie. Though the force does not affect outcomes for purely self-interested bargainers, its relevance for realistic preferences is indicated by the very different experimental results for dictator and ultimatum games; observed offers are much higher in ultimatum games. While both games have $\rho = 0$, the dictator game has $\pi = 1$ and the ultimatum game has $\pi = 0$. In the dictator game, the Proposer can be oblivious to the Decider's sense of fairness; in the ultimatum game, he very much must take it into account.

Because ρ is a measure of how much the Decider is guaranteed without the Proposer's “cooperation,” and π is a measure of how much the Proposer can grab without the Decider's cooperation, in all PDGs $\rho + \pi \leq 1$. When $\rho + \pi = 1$, the scope for bargaining is essentially eliminated, and it is clear that the subgame-perfect equilibrium outcome when the two players are each trying to maximize their money outcome is (π, ρ) . (It will turn out that this split holds when the Decider is motivated by fairness as well.) But when $\rho + \pi < 1$, there is room for bargaining, in the sense that the portion $1 - \rho - \pi$ requires the consent of both parties before it can be allocated.⁵

I now turn to a final characteristic of PDGs that will be of interest. The *starkness* of a PDG is a (very artificial) measurement of a useful concept determining the Decider's ability to harm the Proposer:

Definition 4: Let $r(x) \equiv \sup_{\gamma} \{g_1(1-x, \gamma) \mid g_1(1-x, \gamma) < 1-x\}$. If $g_1(1-x, \gamma) = 1-x$ for all γ , then $r(x) = 0$. The *starkness* σ of a PDG is $\sigma \equiv \text{Min}_{x \in [0,1]} \frac{1-x-r(x)}{1-x}$

The measure $r(x)$ is the largest amount the Decider can give to the Proposer who has just offered x to her *without fully accepting the offer*. From this, we can derive a measure σ of the minimum the Proposer can be punished (if he's punished at all) if he makes an offer x . Starkness differentiates the ultimatum and dictator games from more “incremental” games, which allow some punishment short of destroying the whole pie. The ultimatum game has starkness $\sigma = 1$, because the Proposer will get nothing out of any offer that is not fully accepted. By the convention incorporated into the definition, $\sigma = 1$ for the dictator game as well. In both games, the Decider doesn't really ask herself “how much” she will punish the Proposer—either she punishes him severely, or not at all. While this measure is somewhat artificial, it will provide some intuition of a factor leading to inefficiencies in games.

When $\rho = 0$ and $\pi = 0$, PDGs allow the Decider the same choices as in the Ultimatum game (to accept the offer as is, or to reject it completely) but also to do something in between. I now present two examples of PDGs with $\rho=0$

⁵ All these points also hint at a simple fact: π and ρ are, respectively, the Proposer's and the Decider's minmax payoffs in every PDG, if players' utility payoffs were assumed to be equal to their money payoffs.

and $\pi=0$. The first example is *The Squishy Game*, which provides an “incremental” alternative to the all-or-nothing nature of the ultimatum game:

The Squishy Game: $Y = 0$ and for all γ , $(g_1, g_2) = ((1-\gamma)(1-x), (1-\gamma)x)$.

In the Squishy Game, after the Proposer makes a proposal for splitting the pie, the Decider must accept the proposed proportions, but she gets to decide how big the pie is. By choosing $\gamma = 0$, she can leave it as is: This is like saying “Yes” in the ultimatum game. By choosing $\gamma = 1$, she can shrink the pie to nothing: This is like saying “No” in the ultimatum game. But she can *also* choose to shrink the pie by any amount in between. For instance, she can chop the pie in half, giving both herself and the Proposer one half of what was originally proposed.

The Squishy Game can be interpreted as an attempt to cram into the PDG framework a protracted bargaining process with lopsided bargaining power. In particular, as a variant of the well-known alternating-offer bargaining game (Rubinstein (1982)), consider parties bargaining over a pie where delay in reaching an agreement is costly. Suppose only one player (the Proposer) can make offers, and the other (the Decider) says “Yes” or “No” to these offers. Bargaining continues until the Decider says “Yes” to an offer. Each player dislikes delay.

The strategy spaces in such a game are large and complex, and include strategies that cannot be replicated in the Squishy Game. The Squishy Game essentially allows only strategies where the Proposer chooses some offer and proposes it each period forever, and the Decider chooses when to say yes to this offer. The Squishy Game captures how protracted bargaining provides an opportunity for a player to sacrifice a little bit to hurt the other player by a little bit, rather than presuming that a player who has little bargaining power can only obstruct agreement on an all-or-nothing basis. In this sense, it indicates that the ultimatum game is a misleading proxy for all games of lopsided bargaining power.

I now turn to another simple example of a PDG, where, as with the ultimatum game, the Decider has only two choices. Here, however, the Decider cannot necessarily drive the Proposer's payoffs down to zero. While the game can represent many different situations, I call it *the Sabotage Game* because it captures in simple form a situation where one party unilaterally chooses a split of a surplus, but then the other party can choose a costly act of sabotage that hurts both parties.

The Sabotage Game: $Y = 0$, and there exist parameters $\alpha, \beta > 0$ where, for all $\gamma > 0$, $(g_1, g_2) = (1-x-\text{Min}[1-x, \alpha], x-\text{Min}[x, \beta])$.

In the Sabotage Game, the Proposer makes an offer, $(1-x, x)$, to the Decider on how to split the pie. If the Decider accepts the offer without complication, they each get this split. But the Decider can also choose to sabotage, and cause harm $\alpha > 0$ to the Proposer at cost $\beta > 0$.⁶

These games are different functional forms of Proposer-Decider games. They all have $\rho = 0$, so conventional theory predicts in each of them that the Proposer gets all the surplus.

⁶ The specification I have given assumes that neither player's payoff can go below zero. This is to make the Sabotage Game fit within the definition of PDGs, and is not qualitatively important.

3. Equity, Fairness, and Revenge

The definition of PDGs given in the previous section refers to the “physical” game, and specifies “payoffs” in terms of money outcomes. Proper analysis, however, requires us to specify outcomes in terms of players’ utilities, and a premise of this paper is that the players’ payoffs are not necessarily the same as the money outcomes. For simplicity, however, I assume that the Proposer’s preferences are straightforwardly pecuniary: $U_1(g_1, g_2, x, Y) \equiv g_1$.⁷ I concentrate instead on delineating the Decider’s tastes, and in particular allow the Decider to have a taste for lowering the Proposer’s payoffs whenever the Proposer has proposed an unfair split. The Decider’s payoffs can depend on all the relevant variables and parameters—on their monetary payoffs, g_1 and g_2 , the offer, x , and the minimum allowable offer, Y .

Definition 5: The Decider’s preferences $U_2(g_1, g_2, x, Y)$ are *self interested and fairness motivated* (she has SIFM preferences) if there exists a function $E(Y)$ such that:

- a) U_2 is continuous in all variables;
- b) U_2 is increasing in g_2 ;
- c) When $x \geq E(Y)$, U_2 does not depend on g_1 ;
- d) When $x < E(Y)$, U_2 is decreasing in g_1 for all $g_1 > 1 - E(Y)$;
- e) For all x, Y , the set of indifference curves has slope $dg_2/dg_1 \geq 0$;
- f) There exists $K > 0$ such that, for all x, Y, g_1, g_2 , these indifference curves have slope $dg_2/dg_1 \leq K$;
- g) For all Y and $\epsilon > 0$, there exists $L(Y, \epsilon) > 0$ such that for all $x < E(Y) - \epsilon$ and for all $g_1 > 1 - E(Y) + \epsilon$, the slope of these indifference curves $dg_2/dg_1 \geq L(Y, \epsilon)$.

Condition (a) says that the Decider’s preferences are continuous in all variables, and condition (b) says that she prefers more money to less. The remaining conditions all rely on an “equity function” defined below, which defines (as a function of the parameter Y) the amount of the pie that the Decider feels is “equitable”. Condition (c) says that if the Proposer offers the Decider at least this equitable amount, then the Decider has no desire to hurt the Proposer; condition (d) says that if, however, the Proposer offers the Decider less than the equitable level, the Decider wishes to punish the Proposer, and is willing to sacrifice some (perhaps small) amount of her monetary payoff to do so. Condition (e) (when coupled with (b)) says that the Decider is never willing to sacrifice her monetary payoffs to help the Proposer.⁸ Conditions (f) and (g) put minimal restrictions on how much and how little the Decider is willing to punish the Proposer. Condition (f) merely says that there is *some* upper bound on “the ratio of sacrifice” the Decider is willing to engage in; she is not willing to sacrifice much to harm the Proposer by an arbitrarily small amount. Condition (g) says that for any offer that is strictly less generous to the Decider than the equitable payoff, there exists *some* lower bound on the cost of sacrifice the Decider is willing to pay to hurt the Proposer to the point that the Proposer is no better off than had he made an equitable offer.

⁷ For reasonable ancillary assumptions, I do not believe many qualitative results would change if we assumed that the Proposer himself cared intrinsically about fairness.

⁸ As with the assumption that the Proposer has no direct concern for fairness, I do not believe this assumption about the Decider’s preferences is central to qualitative results.

These preferences correspond loosely to findings in psychological research.⁹ They also correspond to formalizations recently proposed in experimental economics.

By assuming that U_2 depends only on the array (g_1, g_2, x, Y) , I am implicitly ruling out the possibility that the Decider's preferences depend on any features of the PDG except for the parameter Y . This restriction is surely wrong in details, but I feel it does no violence to any important psychological regularities, and is useful for providing simple comparative statics across different bargaining structures.

$E(Y)$ represents the "equity point", above which the Decider considers an offer generous, and therefore has no wish to retaliate, and below which the Decider has the desire to retaliate. The nature of the function $E(Y)$ is, of course, of central importance. Specifying an appropriate $E(Y)$ function involves substantive psychological assumptions about when a person wants to hurt another in a bargaining setting. Bolton (1991) assumes that the Decider feels envy or "inequality aversion"—she doesn't like the idea of walking away from a situation with less money than the Proposer. Thus, choosing $E(Y) = 1/2$ would match Bolton's assumption that when the Proposer is getting more than half the pie, the Decider is willing to sacrifice to lower the Proposer's payoff, and when the Proposer offers more than half the pie to the Decider, the Decider will certainly accept such an offer.

Bolton's model also assumes that U_2 doesn't depend on x or Y . Rabin (1993) proposes a model of reciprocal fairness that assumes a person's taste for hurting another person very much depends on both the behavior of that other person (here, x) and the opportunity set from which that other person chooses his behavior (here, Y). This specification captures the intuition that it is not the inequality of the outcome *per se* that leads Deciders to reject offers in the ultimatum game, but rather the fact that Proposers *choose* to make uneven offers. Consider how people would react if (say) the Proposer offered to keep \$9 for himself and give \$1 to the Decider—but the Decider knew that the rules of the game were that the Proposer was *not allowed* to make any offer of more than \$1 to the Decider. In this context, the Proposer is not being ungenerous by offering the Decider only \$1, and so the Decider would be far less likely to reject the offer than in the ultimatum game. The preferences posited by Bolton would assume that the propensity to turn down (\$9,\$1) offers would be identical in the two games.¹⁰

To capture this, one could assume $E(Y)$ is increasing Y , and that the slope of the indifference curves in g_1 - g_2 space is steeper for lower x .¹¹ Most results of interest will hold for *any* SIFM preferences. Nonetheless, the assumption that the Decider never wishes to punish the Proposer when he offers $x \geq 1/2$ does seem quite realistic, and will prove to be of interest:

Assumption A: For all Y , $E(Y) \leq 1/2$.¹²

⁹ See, e.g., Adams (1963) and Berscheid, Boye, and Walster (1968).

¹⁰ For a test related to the distinction between these two types of fairness, see Blount (1995).

¹¹ Because Rabin's model restrictively assumes that any behavior by another person that involves that person grabbing as much as possible at the expense of the other person is viewed as a mean act, even if "as much as possible" is itself a reasonably fair outcome, technically his model implies $E(Y) > Y$ for all Y . This seems unrealistic in this context, because a split of (.5,.5) in this context seems manifestly fair, and not likely to inspire the ire of the Decider. To capture this, one could very easily modify and extend the model in a variety of ways.

Virtually all the qualitative results presented in the next section do not require Assumption A or anything more than the definition of SIFM preferences; but I shall present several corollaries that reflect the interpretation with this minimal additional assumption.

4. Results

In this section, I consider some general results about outcomes that we can predict given the class of games and class of preferences under consideration. In all results of the paper I specify the unique allocation consistent with a subgame-perfect equilibrium of the game.¹³ Because an offer of $(1-x, x) = (1-\text{Max}[E(Y), Y], \text{Max}[E(Y), Y])$ is always feasible for the Proposer and acceptable to the Decider, it is clear that any equilibrium of a PDG must give at least $1 - E(Y)$ to the Proposer:

Lemma 2: For all PDGs, for all SIFM preferences, the Proposer's equilibrium payoff must be at least $1 - E(Y)$.

Proof: Given the assumptions, the Proposer can get $1 - E(Y)$ by proposing $(1-E(Y), E(Y))$. This clearly establishes a lower bound on his payoffs.

It turns out that, without making stronger assumptions on both the nature of the PDG and the SIFM preferences, Lemma 2 is virtually the *only* conclusion we can reach. Proposition 1 shows that for any SIFM preferences and for any allocation that gives the Proposer at least a payoff of $1 - E(Y)$, there exists a PDG for which this allocation is the equilibrium outcome:

Proposition 1: Consider any Y and any SIFM preferences. Then for every $z_1 \in (\text{Min}[1-E(Y), 1-Y], 1-Y)$ and every $z_2 \in [Y, 1 - z_1]$, there exists a PDG with $\rho = Y$ such that (z_1, z_2) is the unique outcome of PDG.

Proof: Choose any SIFM preferences, and Y , and any (z_1, z_2) meeting the premise. If $E(Y) \leq Y$, then the equilibrium is simply that the Proposer offers $(1-Y, Y)$, and the Decider accepts, so that the Proposition's claim holds in this case. Suppose $E(Y) > Y$. Suppose $z_1 + z_2 = 1$. Then let the PDG be such that $Y = z_2$ and for all $x \geq Y$, for all γ , $g_1(x, \gamma) = 1 - x$ and $g_2(x, \gamma) = x$. (This is a truncated dictator game.) Then clearly the equilibrium will be for the Proposer to offer (z_1, z_2) .

Suppose $z_1 + z_2 < 1$. By Condition (g) of the definition of SIFM preferences we know that:

There exists $x^* > z_2$ such that $U_2(z_1, z_2, x^*, Y) > U_2(1-x^*, x^*, x^*, Y)$; and

There exists $\varepsilon > 0$ such that for all $x \in [Y, 1-z_1]$, $U_2(1-E(Y), x-\varepsilon, x, Y) > U_2(1-x, x, x, Y)$.

Let PDG(ε) be the mechanism that involves a binary choice by the Decider for all x , where:

$$(g_1(x, \gamma), g_2(x, \gamma)) = (1-E(Y), x-\varepsilon) \text{ for all } \gamma > 0 \text{ and all } x \neq x^*$$

$$(g_1(x^*, \gamma), g_2(x^*, \gamma)) = (z_1, z_2) \text{ for all } \gamma > 0$$

¹² The focus here on "1/2 the pie", as opposed to, say 40% or 60% of the pie, while of course important in practical terms, is not central for the qualitative results.

¹³ Technically, I have not made specific assumptions to guarantee uniqueness. While uniqueness is guaranteed for the pure self-interest case, sets of indifference curves for SIFM preferences could precisely correspond to the structure of PDGs to yield multiple equilibria. Such multiplicity won't occur generically in g_1 - g_2 space, so I am ignoring the possibility.

The equilibrium of this game will be that the Proposer offers $(1-x^*, x^*)$, and the Decider responds with $\gamma > 0$, yielding payoffs (z_1, z_2) . Q.E.D.

This proposition says that for any allocation where the Proposer gets more than $1-E(Y)$, there exists a bargaining structure yielding that outcome in the unique equilibrium. Proposition 1 says more than that any offer by the Proposer is possible—it says that any combination of offer by the Proposer and “degree of rejection” by the Decider is possible, so long as it leaves the Proposer with more than $1-E(Y)$. Corollary 1.1 shows that we know *especially* little about outcomes when the Proposer has all the bargaining power (as conventionally conceived):

Corollary 1.1: If $E(0) = 1/2$, then for all $z_1 \in (1/2, 1]$ and $z_2 \in [0, 1-z_1]$, there exists a PDG with $Y = 0$ such that (z_1, z_2) is the unique outcome of PDG.

We know only that the Proposer gets more than half the pie. To make stronger predictions, we need to know more about details of the bargaining environment “classically” deemed not to matter.

Note that any outcome $z_1 + z_2 < 1$ is inefficient. Thus, Proposition 1 says that any SIFM preferences can lead to an inefficient outcome depending on the bargaining structure. I shall use as a definition of efficiency the notion of “wealth maximization,” considering an outcome efficient if the entire pie is split. This definition is problematic, especially because the very premise of that result is that the Decider’s payoff includes a non-monetary component. Nonetheless, it is of some use for characterizing outcomes. Moreover, if neither party finds it pleasant when one person has taken revenge on another person whom she thinks has behaved unfairly, we might also imagine that the non-pecuniary welfare is also lower when some money is wasted, so that these outcomes are inefficient in a broader sense.

Definition 6: An equilibrium is τ -efficient if it yields payoffs $g_1 + g_2 \geq \tau$.

Proposition 2 formalizes a trivial result that a lower bound on efficiency is determined by what the Proposer can guarantee for himself:

Proposition 2: For all SIFM preferences, and all PDGs with force π , the outcome will be π -efficient.

Proof: By definition of the force, there exists x such that $g_1(x, \gamma) \geq \pi$ for all γ . Therefore, the Proposer will choose an offer that yields him at least this payoff.

An immediate corollary of interest is that if $\pi = 1$, we get full efficiency. That is, if the Proposer has sufficient power to force an inequitable solution irrespective of what the Decider wants, then we get full efficiency. Thus, for instance, if a firm cannot be harmed by retaliation, we will always get an efficient outcome.

But perhaps the most important result is the following one:

Proposition 3: For all SIFM preferences and for all $\tau < 1$, there exists an $\varepsilon > 0$ such that the outcomes of all PDGs where $Y > E(Y) - \varepsilon$ are τ -efficient.

Proof: Choose preferences and τ . If $Y \geq E(Y)$, all offers will be accepted, so the result holds. Consider $Y \in (E(Y) - \varepsilon, E(Y))$. By Lemma 2, we know that the Proposer will never make an offer yielding a response giving him $g_1 < 1 - E(Y)$. By condition (f) of SIFM preferences, we know that there exists K^* such that there exists $\delta^* > 0$ such that the Decider would never respond to an offer by punishing the Proposer by less than δ^* if it costs her more than $K^*\delta^*$. Letting $\varepsilon = \delta^*$, we know that any equilibrium will be $(1 - (K^* + 1)\varepsilon)$ -efficient. Thus, for any τ , we choose ε such that $1 - (K^* + 1)\varepsilon > \tau$. QED.

Proposition 3 says that if the Proposer simply cannot make low offers to the Decider—if Y is high—we should expect high efficiency. The intuition is that, with any upper limit on the Decider’s willingness to punish the Proposer by only a small amount, any significant inefficiency must hurt the Proposer by some significant amount. But this implies that when the most the Proposer can *try* to grab, $1 - Y$, does not far exceed his “equitable” share, $1 - E(Y)$, he will never make an offer that will be significantly retaliated against; he has little to gain, and lots to lose by trying to be unfair to the Decider.

Proposition 3 argues that when there are inefficiencies, they can be viewed as arising from the asymmetric bargaining power, because the inefficiencies arise only when the Proposer is able to make offers that are far worse than the equitable outcomes. This interpretation is made clearer by Corollary 3.2, which adds Assumption A:

Corollary 3.1: For all SIFM preferences, and for all $\tau < 1$, there exists an $\varepsilon > 0$ such that the outcomes of all PDGs where $\rho > 1 - \varepsilon$ are τ -efficient.

Corollary 3.2: For all SIFM preferences meeting Assumption A, for all $\tau < 1$, there exists an $\varepsilon > 0$ such that the outcomes of all PDGs where $\rho > 1/2 - \varepsilon$ are τ -efficient.

Corollary 3.2 says that full efficiency is guaranteed if $\rho = .5$. We can of course also say exactly what the division will be: $(1/2, 1/2)$.

Taken together, Propositions 2 and 3 say that if we care about wealth maximization, then we may prefer bargaining environments where ρ is close to one half, so that bargaining power is balanced, or, in lieu of balanced bargaining power, we prefer the strong party to be protected from retaliation: $\pi = 1$.

Neither Proposition 2 nor 3 indicates that the ultimatum game will yield an efficient outcome under full information. This, rather, is implied by the starkness of the ultimatum game discussed in Section 2. The following proposition formalizes the assertion that the outcome in the ultimatum game will be efficient irrespective of details of preferences:

Proposition 4: For all SIFM preferences, for all Y , the outcome in every PDG with starkness $\sigma > E(Y)$ will be fully efficient.

Proof: Because the Proposer can guarantee himself $1 - E(Y)$ by offering $(1 - E(Y), E(Y))$, and because any rejected offer will yield him $g_1 < 1 - Y - \sigma < 1 - E(Y)$, he will never make an offer that will be rejected. Q.E.D.

Corollary 4.1: For all SIFM Preferences meeting Assumption A, the outcome in every PDG with starkness $\sigma > 1/2$ will be fully efficient.

Proposition 4 captures why both the ultimatum game and dictator game are unusually prone to efficiency: The dictator game makes the Proposer not worry about his offer being rejected; the ultimatum game provides the Proposer with the most extreme incentives to make sure that whatever proposal he makes will be acceptable to the Decider. Because he has the option of offering the 50/50 fair offer, which he knows will be accepted, the only possible outcome if he fully anticipates the Decider's behavior is for him to make an offer that will be accepted. The fact the Decider has an opportunity to sacrifice a moderate amount to punish the Proposer is what allows inefficiencies. The very starkness that makes the ultimatum and dictator games attractive candidates for testing certain hypothesis makes them misleading in regards to the question of efficiency.

5. An Example

I turn to an example taken from the context of labor economics, where a firm offers a wage to a potential employee, and then the employee decides whether to accept a job, and whether to be productive for the firm or not.

The Wage-and-Work Game: $Y = 0$ and for all $\gamma < \text{Min}[1-x, \sqrt{x}]$, $(g_1, g_2) = (1-x-\gamma, x-\gamma^2)$. For $\gamma \geq \text{Min}[1-x, \sqrt{x}]$, $(g_1, g_2) = (0, 0)$.

I translate the variables of the above definition to yield a more desirable interpretation. Let $e = (1 - \gamma)$. Then the Proposer offers $w \leq 1$, and the Decider chooses $e \in [0, 1]$, yielding $(g_1, g_2) = (e-w, w - (1-e)^2)$. Specifically, the Capitalist proposes to hire the Worker at fixed wage w . The Worker can either reject the offer, or accept the offer. If she accepts the offer, she must then choose "endeavor level" $e \in [0, +\infty)$.¹⁴ The Capitalist's profits are $\pi(e, w) \equiv e - w$, where $e = w = 0$ if the Worker turns down the Capitalist's job offer. The Worker's "material well-being" is 0 if she rejects the offer, and is $y(e, w) \equiv w - (e-1)^2$ if she accepts the job. This functional form represents the fact that the Worker's self interest lies in putting in endeavor/effort level $e = 1$ after accepting the job. This variable has different interpretations. One is that it is an effort level, but where the higher the effort the Worker puts in, the more likely she is to build a good reputation. Given career and employability goals, up to a certain level, high effort pays off for her. Another interpretation is that $e = 1$ involves literally no effort; the organization is set up such that this comes naturally, and it would be more effort to lower the Capitalist's profits than to keep them this high. One might even call increasing $1 - e$ to be sabotage. Of special note is that starting from the efficient level, $e = 1$, a small reduction in the Worker's endeavor level will have only second-order costs to her but will inflict first-order costs on the Capitalist.

¹⁴ Note that only values $e < 1$ are candidates for equilibrium outcomes here. Because I have developed the model here to focus on a particular set of issues, the assumptions have ruled out the possibility of a $e > 1$ equilibrium, where the worker sacrifices his well-being to help the Capitalist.

To analyze the Wage-and-Work Game, I assume that, instead of simply preferring to maximize her “material well-being,” the Worker wishes, in addition, to minimize the level of inequity in the distribution of her production. In particular, assume that

$$U(e,w) \equiv w - (1-e)^2 - d\text{Max}[0, \pi(e,w) - y(e,w)],$$

where $d > 0$ is a parameter measuring how distasteful the Worker finds it have the Capitalist get more from her endeavor to be productive than she does herself.

What wage offers will the Worker accept, and what endeavor level e will she choose in response to such wages? What wage will the Capitalist offer? By solving the first-order conditions, if the Capitalist offers the Worker a wage of $w < 1/2$, then the Worker (if she accepts the job) will choose effort $e(w) = 1 - d/(2(1+d))$. If d is small enough, then the Worker will indeed take the job. If $w \geq 1/2$, then the Worker will accept the job, and choose $e = 1$, and there will be full efficiency.

Thus, the Capitalist's choice is clear: She can choose either $w = 1/2$, guaranteeing that she'll get profits of $1/2$, or she can offer the Worker a low wage that will cause the Worker to accept the job but choose $e < 1$. The optimal low wage for the Capitalist solves:

$$\begin{aligned} &\text{Max } e(w) - w \\ &\text{s.t. } y(e(w),w) - d\text{Max}[0, \pi(e(w),w) - y(e(w),w)] > 0. \end{aligned}$$

Because $e(w) = 1 - d/(2(1+d))$ is independent of w , this constrained-maximization problem is solved by choosing the lowest w such that the constraint holds. This turns out to be $w = d(3d+4)/(4(1+d)(1+2d))$, yielding the Capitalist profits of $\pi(e(w),w) = 1 - d(d-2)/(4(1+d)(1+2d))$. As $d \rightarrow 0$, $\pi(e(w),w) \rightarrow 1 > 1/2$. This implies that for d sufficiently small, the Capitalist will choose to pay $w < 1/2$, inducing $e < 1$. Therefore, for this range of values of d , there will be inefficient production.

6. Discussion and Conclusion

I conclude by conjecturing some of the ways the lessons of this paper are likely to extend beyond the narrow framework presented, and by highlighting several caveats to the conclusions I've reached. The most obvious way to generalize the model is to consider bargaining structures that consist of more than two stages. Allowing preferences to depend on the other player's behavior would complicate the analysis, because the distinction between what is nice and mean behavior will be far more complicated than in PDGs, where it is natural to interpret lower offers as meaner. Also, any analogue to Proposition 3 would be more complicated, because there may not be any simple way to interpret bargaining power, ρ , as there was in PDGs (namely, $\rho = Y$).

Another direction for extending the model to check its robustness is perhaps more urgent: I have assumed throughout both the discussion and formal analysis that the exact preferences of the Decider are common knowledge. This allowed me to emphasize that inefficiencies did not derive from incomplete information, which we know can

often lead to inefficiencies in bargaining settings. But the results may be misleading, and incomplete information is likely to moderate the basic arguments of the paper. For instance, under complete information, the “starkness” of the ultimatum game guarantees its efficiency, because the Proposer will always avoid the Decider’s retaliation when he correctly predicts it. But, realistically, the Decider’s “threshold” is often *not* common knowledge; self-interested Proposers will accept some probability of rejection. Hence, ultimatum games will generate inefficient punishments, just as alternative bargaining structures do. In a world of uncertainty, where retaliation happens with positive probability, the more moderate nature of punishments in alternative bargaining structures might lead to greater efficiency than in ultimatum games.

A final caveat is entirely speculative, but it highlights the limits of both structured laboratory experiments and simple formal models of preferences of the form I have posited. Bargaining is a process, rather than simply a mapping from some social opportunity set to an outcome. During this process, people posture, fight, and argue, often generating hostility. It may be that, on average, more equal bargaining power leads to more bargaining tensions, which in turn leads to more hostility and less cooperation. When bargaining power is unequal, it may be that such tensions are mitigated. I think this hypothesis is wrong in its simple form, but there may be reason to believe that in some settings more total surplus will be dissipated when both parties feel they can obtain sizable shares of the surplus through aggressive bargaining.

There is one direction that I suspect the results are likely to extend, and be strengthened: I don’t feel that the qualitative results of this paper rely on it being costly to the Decider to harm the Proposer. “Retaliation” may instead involve her withholding (costly) cooperation from the Proposer when he has mistreated her. Akerlof (1982), for instance, argues that labor effort can be seen in part as such a “gift exchange” relationship. The model of Section 5 could readily be modified to provide such a model. If we define full efficiency as the maximum payoffs possible when parties make the type of small sacrifices people make all the time to cooperate with one another, then the arguments of this paper about when inefficiencies arise are likely to hold even more generally.

REFERENCES

- Adams, J. S. (1963). “Wage Inequities, Productivity and Work Quality.” *Industrial Relations* 3: 9-16.
- Akerlof, G. A. (1982). “Labor Contracts as Partial Gift Exchange.” *Quarterly Journal of Economics* 97(November): 543-569.
- Akerlof, G. A. (1984). “Gift Exchange and Efficiency-Wage Theory: Four Views.” *American Economic Review* 74(2): 79-83.
- Akerlof, G. A. and J. L. Yellen (1990). “The Fair Wage-Effort Hypothesis and Unemployment.” *Quarterly Journal of Economics* 105(2): 255-283.
- Berscheid, E., D. Boye and E. Walster (1968). “Retaliation as a Means of Restoring Equity.” *Journal of Personality & Social Psychology* 10(4): 370-376.

- Blount, S. (1995). "When Social Outcomes Aren't Fair: The Effect of Causal Attributions on Preferences." *Organizational Behavior and Human Decision Processes* 63(2): 131-144.
- Bolton, G. E. (1991). "A Comparative Model of Bargaining: Theory and Evidence." *The American Economic Review* 81(5): 1096-1136.
- Camerer, C. and R. Thaler (1995). "Anomalies: Ultimatums, Dictators and Manners." *Journal of Economic Perspectives* 9(2): 209-219.
- Guth, W., R. Schmittberger and B. Schwarz (1982). "An Experimental Analysis of Ultimatum Bargaining." *Journal of Economic Behavior and Organization* 3: 367-388.
- Guth, W. and R. Tietz (1990). "Ultimatum Bargaining Behavior: A Survey and Comparison of Experimental Results." *Journal of Economic Psychology* 11(3): 417-449.
- Mars, G. (1982). *Cheats at Work: An Anthropology of Workplace Crime*, George Allen & Unwin.
- Rabin, M. (1993). "Incorporating Fairness into Game Theory and Economics," *American Economic Review* 83(5): 1281-1302.
- Roth, A. E. (1995). *Bargaining Experiments*. *Handbook of Experimental Economics*. J. Kagel and A. Roth. Princeton, Princeton University Press: 253-348.
- Rubinstein, A. (1982). "Perfect Equilibrium in a Bargaining Model," *Econometrica* 50(1): 97-109.
- Thaler, R. H. (1988). "Anomalies: The Ultimatum Game." *Journal of Economic Perspectives* 2(4): 195-206.