Department of Economics
ISSN number 1441-5429
Discussion number 02/19

Are Efficient Bargaining Power Disparities Unfair? An Experimental Test

Aaron Nicholas[1] and Birendra Rai[2]

Abstract:
A key question in labor and contract law is when does bargaining power disparity become too large to be considered `impermissible'? It has largely been debated from the potentially conflicting perspectives of efficiency and fairness. These debates exhibit the intuitively plausible but empirically untested presumption that efficient bargaining power disparities can be unfair. The paper focuses on ex-post bargaining between agents locked in a relationship without a complete contract wherein surplus may ultimately be realized with or without mutual consent. We propose a consent-based definition to categorize a bargaining power disparity as either efficient or inefficient by treating surplus realized without mutual consent as an imperfect substitute for surplus realized with mutual consent. In order to categorize a power disparity as either fair or unfair, we draw upon some legal doctrines to propose a two-sided definition that accounts for the perspectives of both the weaker and the stronger bargaining parties. The experiment provides no robust evidence to support the presumption that economically efficient power disparities can be unfair.

Keywords: Bargaining power, consent, efficiency, fairness, law, contract, experiment

JEL Codes: K0, C72, C91, D63

Acknowledgement: We thank Tim Cason, Marco Castillo, Uwe Dulleck, Patrick Francois, Jacob Goeree, John List, Ragan Petrie, Daniela Puzzello, Joyce Sadka, Vernon Smith, Tom Wilkening, Bart Wilson, numerous colleagues in our departments, seminar audiences at Paris School of Economics, University of Exter, and participants at the 12th Australia New Zealand Workshop on Experimental Economics. We are particularly grateful to Toby Handfield, Peter Lambert, and Sarah Meehan for detailed comments. The research was funded by a grant from the Department of Economics at Monash university. The usual caveat applies.

[1] Department of Economics, Deakin University, Burwood 3125, Australia.
[2] Department of Economics, Monash University, Clayton 3800, Australia.

1. Introduction

When does bargaining power disparity become too much to invite intervention? A voluntary exchange between agents that does not create externalities would be Pareto improving, and thus intervention is difficult to justify on grounds of efficiency regardless of the level of bargaining power disparity between agents. Efficient power disparities may nonetheless seem unfair especially when the power disparity between agents is sufficiently large. Most legal policy debates have to confront this potential tension between economic efficiency and fairness (Kaplow and Shavell, 2001). Cooter and Rubinfeld (1989), for example, highlight that "legal policy has traditionally been evaluated by standards of fairness ... efficiency is more controversial as a goal for law as opposed to markets."

While the difficulty in judging how large a power disparity must be to be considered unfair needs little elaboration, it is worth noting that the efficiency perspective faces the challenge of establishing that an interaction is indeed mutually consensual (Llewellyn, 1960; Bowles and Gintis, 1992; Craswell, 1993; Posner, 1995). This is especially the case when agents are locked in a relationship without a complete contract. Such contexts involve "a nonconsensual penumbra around a consensual core" (Fried, 2015 pp. 72) and create scope for opportunism, i.e., for the stronger party to profit without the consent of the weaker party. While this point has been recognized in the economics literature (Williamson, 1985; Basu, 2007; Piccione and Rubinstein, 2008; Acemoglu and Wolitzky, 2011), how surplus realized without mutual consent should enter the calculus of efficiency remains unexplored.

Our goal is to experimentally test the presumption that economically efficient bargaining power disparities can be unfair. The paper takes a revealed preference approach and proposes two non-parametric operational definitions that respectively categorize bargaining power disparities as efficient or inefficient and fair or unfair on the basis of agents' behavior. The basic premise of our work is that 'size of the pie' is a measure of *accounting* efficiency (Posner, 1980; Schelling, 1981; Fogel, 1990). Assuming mutual consent is central to the idea of *economic* efficiency, we propose a definition to categorize power disparities as either efficient or inefficient that accounts for *how* the pie is realized: with or without mutual consent. With respect to fairness, we draw upon existing legal doctrines to propose a *two-sided* definition that seeks to account for the perspectives of both the weaker and the stronger bargaining parties but does not rely on an external observer's perceptions of distributive fairness between the bargaining parties.

To fix ideas, suppose two agents initiate a transaction without a complete contract and negotiate how to share the surplus after the surplus $\pi > 0$ has been generated. Let the game $\mathcal{U}(x)$, a simple variant of the ultimatum game, represent a caricature of the terminal stages of this *ex-post* negotiation (Figure 1). The proposer P makes a take-it-or-leave-it offer to the responder R. The realized material surplus at any $x \in [0, \pi]$ is π if the proposer's offer to divide π garners the consent of the responder. In the absence of such mutual consent, the realized material surplus is x and the proposer obtains all of it. We interpret rejection and the accompanying lack of mutual consent as the beginning of a dispute between the parties. We are agnostic about how payoffs are enforced in

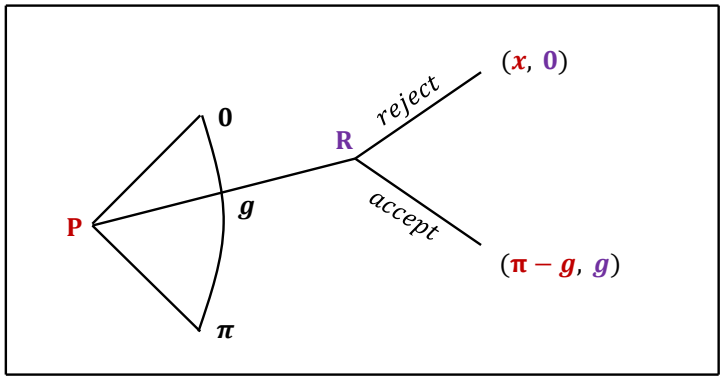

Figure 1: *The ex-post bargain $\mathcal{U}(x)$ between agents locked in a relationship.* The realized material surplus is either π or x depending upon whether or not P's offer garners the consent of R. Our interpretation of $\mathcal{U}(x)$ as an *ex-post* bargain differs from its interpretation as an *ex-ante* bargain wherein the latter assumes that (i) acceptance by R sets the terms at which P and R get locked-in to generate and share π, (ii) rejection by R implies the agents do not engage in the transaction, and (iii) the payoff of x to P following rejection by R is derived from an interaction with and the consent of some hypothetical third party.

the absence of mutual consent. It may or may not involve third-parties, such as the law. Different values of x may be interpreted as different legal regimes that serve as the background for negotiation between the involved parties (Hayek, 1960; Mnookin and Kornhauser, 1979). In an experiment, the experimenter serves as the enforcer.

$\mathcal{U}(0)$ is the standard ultimatum game where no surplus can be realized without mutual consent. Surplus can be realized with or without mutual consent when $x > 0$. The ability of the responder to materially punish the proposer by rejecting what she considers an unfair offer declines with an increase in x. Following the existing literature we therefore interpret an increase in x as an increase in bargaining power disparity in favor of the proposer (Binmore et al., 1991; Anbarci and Feltovich, 2013). Our interest lies in developing a way to categorize each x as efficient or inefficient and fair or unfair; and, assess whether any efficient power disparity is unfair.

To see the motivation behind our efficiency definition, imagine an economy where pairs of agents bargain as per $\mathcal{U}(x)$ and a planner must choose the level of power disparity x. Suppose the planner is guided solely by efficiency concerns and anticipates the aggregate rate of agreements across all bargaining pairs to decrease with an increase in x. If the planner employs *total* surplus realized by the bargainers as the operational measure of efficiency, then she would choose the highest level of power disparity which guarantees 'no money will be left on the table'. In contrast, if she uses *agreement* surplus (i.e., surplus realized with mutual consent) as the operational measure of efficiency, then she would choose the lowest level of power disparity.

The total surplus measure treats disagreement surplus (i.e., surplus realized without mutual consent) as a perfect substitute for surplus realized via agreements. It is therefore best viewed as a measure of accounting efficiency rather than economic efficiency.[1] The agreement surplus measure

[1]If agents' preferences are sensitive to how surplus is realized, then Pareto efficiency or social welfare maximization

effectively treats disagreement surplus as a *bad* at all levels of power disparity. While it is unclear whether disagreement surplus should be considered a bad at every power disparity, assuming that it is never a bad seems normatively indefensible. It would imply that any shortfall in agreement surplus can be compensated by some suitably large increase in disagreement surplus. One could then justify exploitative institutions such as slavery on grounds of economic efficiency.

We propose a *consent-based* definition to categorize power disparities as economically efficient or inefficient. The basic idea rests on comparing the surplus vectors at any pair of power disparities, where the surplus vector at a power disparity refers to the agreement surplus and disagreement surplus at that power disparity generated across all bargaining pairs. Our definition treats disagreement surplus as (a) an imperfect substitute for agreement surplus under all circumstances, and (b) a bad under some circumstances. Suppose total surplus – the sum of agreement and disagreement surplus – is lower at power disparity x than at y. We define x to be *objectively* inefficient relative to y if agreement surplus is relatively lower and disagreement surplus is relatively higher at x. Further, if and when we are willing to treat disagreement surplus as a bad, then y can be considered *subjectively* inefficient relative to x in case the relatively higher total surplus at y is entirely due to disagreement surplus being relatively higher at y. We define a power disparity to be inefficient if it is objectively or subjectively inefficient relative to some power disparity.

Turning to fairness, notice that there is no existing definition of an (un)fair power disparity that we could readily utilize. The bargaining literature in economics invariably discusses fairness within particular interactions, and while doing so focuses on the perspective of the weaker party. For instance, upon observing a rejection by a responder in the ultimatum game the analyst concludes that the responder finds the offer unfair; but, there is no discussion as to whether there is any way to justify to the proposer that his offer is indeed unfair. The law, in contrast, makes an attempt to do so by asking what would a *reasonable person* in the role of the stronger party have done. The intention being to test the validity of the stronger party's (implicit or explicit) defense that his behavior is fair because reasonable others would behave similarly under the given circumstances.

We say a power disparity is not *dispute-proof* if average offer by proposers is lower than the average minimum acceptable offer (MAO) of responders at that power disparity. Dispute-proofness simply transposes to the aggregate level the standard interpretation that a particular responder's MAO is the minimum offer that she personally considers fair. Next, we say a power disparity is not *selfish-proof* if average offer by self-regarding proposers is lower than average offer by non-self-regarding proposers.[2]. We define a power disparity to be unfair if it is neither dispute-proof nor selfish-proof.

Based on assumptions consistent with existing empirical evidence we theoretically argue that sufficiently low power disparities are likely to be fair but sufficiently high power disparities may be fair or unfair. However, no general theoretical predictions can be made about which power dispari-

---

would be misleading if the analyst assumes agents' utility functions are defined solely over outcomes.

[2]Self-regarding proposers are defined as those who care solely about material payoffs to themselves. Section 2 explains why we treat non-self-regarding proposers as the 'reasonable' proposers.

ties are likely to be inefficient. Depending upon how agreement rates vary across power disparities, the set of inefficient power disparities may be empty or non-empty, connected or disconnected, and involve low or high levels of power disparity. In general, whether efficient power disparities are unfair is sensitive to how agreement rates vary across power disparities.

Our first treatment is based on $\mathcal{U}(x)$. Subjects are assigned to the fixed role of either a proposer or a responder. Proposers report their offers out of π and responders report their minimum acceptable offer out of π at multiple values of x. The second treatment is based on a game where the first mover can *choose* either (i) to take an outside option, or (ii) to enter $\mathcal{U}(x)$ knowing that the pie to be bargained over will be bigger than her outside option but she will be the weaker bargaining party (i.e., the responder) upon entry.

Notwithstanding the theoretical ambiguity, we find a remarkable qualitative similarity in the evaluation of power disparities from the fairness and efficiency perspectives in both treatments. Sufficiently low power disparities are fair and efficient whereas sufficiently high power disparities are unfair and inefficient. Perhaps somewhat surprisingly, we find no robust evidence to support the intuitively plausible presumption that some economically efficient power disparities can be unfair.

Many legal changes have been rationalized partly as attempts to correct for unfairness by curbing systematic bargaining power disparities between the relevant classes of actors. For example, Schwab (2017) and Davidov (2016) highlight that much of labor law throughout the world is still explained primarily with reference to systematic bargaining power disparities. Present day contract law contains numerous doctrines that permit courts to intervene in private contracts between parties with asymmetric bargaining power on grounds of fairness (Scott, 2004). Further, the notion of bargaining power disparity has been used to rationalize several structural transitions in law: the shift from litigation to regulation in the Gilded Age (Glaeser and Shelifer, 2003), the choice between injunctions and compensations in securing property across space and time (Glaeser et al., 2016), and even the carving out of labor law from contract law (Unger, 1983).

Our main contribution lies in offering a revealed preference approach to evaluate bargaining power disparities from the potentially conflicting perspectives of efficiency and fairness. The observed close correspondence between the two perspectives hints towards the possibility that economic efficiency may subsume fairness considerations in the evaluation of power disparities. We arrive at this conclusion in an abstract set-up using novel operational definitions – consent-based efficiency and two-sided fairness – that seek to better approximate the core principle behind economic efficiency and guard against the common critiques in conceptualizing fairness.

We conclude by relating our work to the broader debate about fairness versus efficiency in legal scholarship. We also draw upon some related existing experimental findings to argue that two aspects related to the notion of power disparity – its source and the stage of the relationship at which it is exercised – may be crucial in evaluating power disparities. Together with our findings they hint towards a behavioral rationale for (a) delineating the role of different laws that seek to regulate bargaining power disparities and (b) why a high power disparity, in and of itself, is typically regarded insufficient to warrant intervention.

2. Framework and definitions

Consider the game $\mathcal{U}(x)$ as described in the previous section and illustrated in Figure 1. As x increases, the 'structural' bargaining power disparity between the agents increases in favor of the proposer. Assuming $x \in [0, \pi]$ reflects that available surplus cannot increase following a disagreement. Normalizing the responder's payoffs to zero following a disagreement plays no substantive role in our analysis. The qualifier 'structural' highlights that x is independent of agents' preferences and beliefs.[3] No level of power disparity is labeled unfair or inefficient *a priori*. The goal is to make these judgements on the basis of agents' behavior.

The existing literature seems to interpret a game like $\mathcal{U}(x)$ as an *ex-ante* bargain. Under the ex-ante interpretation, the outcome of the bargain determines whether or not the agents initiate their transaction. Specifically, the transaction does not begin if P's offer fails to garner R's consent regarding the division of the surplus π that would be generated if they were to transact. If P's offer is rejected by R, then the payoff of x obtained by P is assumed to represent what he gets by interacting with, and the consent of, some hypothetical third party R_h. It is worth stressing that the ex-ante interpretation of $\mathcal{U}(x)$ allows the analyst to implicitly *assume* that all surplus is realized with mutual consent between some pair of agents – either real or hypothetical. We are not aware of any study that makes this point explicitly. Nonetheless, it helps understand the normative justification for using total realized surplus as the operational measure of efficiency in the analysis of bargaining games, as is the case in the existing literature (see Hennig-Schmidt et al. (2018) and the references therein).

The present paper, in contrast, interprets $\mathcal{U}(x)$ as an ex-post bargain between agents locked in an interaction without a complete contract. Lock-in with incomplete contracts is pervasive in social and economic relationships due to the "fundamental transformation" that occurs when interactions that start in a market ultimately morph into a bilateral relationship (Williamson, 1985). It creates scope for surplus to be ultimately realized without mutual consent between the parties. Yet, the lock-in under $\mathcal{U}(x)$ is unlike that experienced by the weaker party in a transparently coercive interaction – your life or your wallet – where the weaker party is forced to choose between two of her *own* entitlements (Epstein, 1975). Consequently, we seek a way to define inefficient power disparities that accounts for *how* the surplus is realized.[4]

[3] It is helpful to distinguish between having power versus exploiting it. A relatively higher x implies the proposer has relatively greater bargaining power; whether and how successfully he exploits it will depend on the preferences and beliefs of the agents.

[4] One may argue that the ex-post bargaining as per $\mathcal{U}(x)$ is likely preceded by mutual consent between the agents to *initiate* the transaction, and question the need to distinguish between surplus realized with and without mutual consent. Specifically, suppose two agents initiate a transaction with mutual consent but without a complete contract under common knowledge of the set of potential outcomes. Does consent to initiate a transaction, in and of itself, imply mutual consent towards any outcome that eventuates? The current legal answer is 'no'. Any party can withdraw consent at any time. Of course, depending upon the context, the withdrawing party may have to compensate the other (e.g., several provisions such as reliance and expectation damages in contract law).

For expositional clarity it is helpful to bear in mind a brief outline of our experimental design. Consider an experiment involving 2N subjects, where N subjects are randomly assigned to act as proposers and the remaining N as responders. Subjects make decisions in multiple games that differ only in the value of x. Each proposer reports his offer at each x. Responders report their minimum acceptable offer (henceforth, MAO) at each x without any information about proposers' offers. The definitions for (in)efficient and (un)fair power disparities will be based on the observed distributions of offers and MAOs at different levels of power disparity.

2.1. Defining efficient and inefficient power disparities

Consider a proposer $p \in P$ and a power disparity $x \in [0, \pi]$. Suppose the proposer offers $g^p(x) \in [0, \pi]$ in $\mathcal{U}(x)$. Given the design outlined above, the probability of acceptance for this proposer's offer, denoted by $\alpha^p(x)$, is simply the proportion of responders whose MAO in $\mathcal{U}(x)$ is no more than $g^p(x)$. Suppose each proposer is equally likely to be matched with any responder. The expected *total* surplus across all potential bargaining pairs involving proposer p will be

$$S_t^p(x) \;=\; \alpha^p(x)\pi \;+\; (1-\alpha^p(x))x \;=\; S_a^p(x) \;+\; S_d^p(x).$$

Let $\alpha(x) \;=\; \frac{1}{N}\sum_{p\in P}\; \alpha^p(x)$ denote the expected aggregate agreement rate at x (normalized to lie between zero and one). If agreement rate at power disparity $x \in [0, \pi]$ is $\alpha(x)$, then

- the *agreement surplus* at x will be $S_a(x) = \alpha(x)\pi,$ with $S_a(x) \in [0, \pi]$
- the *disagreement surplus* at x will be $S_d(x) = (1-\alpha(x))x,$ with $S_d(x) \in [0, x]$.
- the *total surplus* at x will be $S_t(x) = S_a(x) + S_d(x),$ with $S_t(x) \in [x, \pi]$.

Consider a power disparity x with surplus vector $S(x) = (S_a(x), S_d(x))$, such that the total surplus at x is $S_t(x) = S_a(x) + S_d(x)$. Let the total surplus at some power disparity $y \neq x$ be $S_t(y) \geq S_t(x)$. The relatively greater total surplus at y may be driven (I) *entirely* by the relatively greater *agreement* surplus at y, or (II) *entirely* by the relative greater *disagreement* surplus at y, or (III) *partly* by both.

Imagine a planner who wants to choose between power disparities x and y on grounds of economic efficiency. If the planner believes mutual consent is central to any reasonable conceptualization of economic efficiency, then she would treat disagreement surplus as worse than a perfect substitute for agreement surplus. Case I poses no dilemma to such a planner and she would choose y over x without any reservations. Case II presents a dilemma because choosing y over x will (a) increase total surplus, but (b) this increase would be achieved entirely via an increase in disagreement surplus. In order to deal with this case the planner would need to first resolve whether and when disagreement surplus should be considered a 'bad.' Case III presents a slightly different dilemma because choosing y over x would not only increase agreement surplus but also disagreement surplus.

Definition $\mathbf{D_E}$. *Consider a power disparity $x \in [0, \pi]$.*

- *$x \in [0, \pi]$ is objectively inefficient relative to power disparity $y \neq x$ if*

$$S_a(y) > S_a(x),\ S_d(y) \leq S_d(x),\ \text{and}\ S_t(y) \geq S_t(x).$$

- *$x \in [0, \pi]$ with $S_d(x) > S_a(x)$ is subjectively inefficient relative to y if*

$$S_a(y) \geq S_a(x),\ S_d(y) < S_d(x),\ \text{and}\ S_t(y) < S_t(x).$$

- *x is inefficient if it is objectively or subjectively inefficient relative to some $y \neq x$.*
- *x is efficient if it is not inefficient.*

$\mathbf{D_E}$ considers a power disparity efficient unless it can be proven inefficient relative to some other power disparity. It is a non-parametric operational definition that does not require information about agents' preferences. Instead, it utilizes information about the size of realized surplus and how the surplus is realized. Among two power disparities, objective inefficiency recommends choosing the power disparity with greater total surplus if doing so would strictly increase agreement surplus and weakly decrease disagreement surplus. For instance, in Figure 2, the power disparity x_o with surplus vector $S(x_o)$ is objectively inefficient relative to any power disparity such as x_1 whose surplus vector $S(x_1)$ lies in Region I. A planner who violates the recommendation by objective inefficiency reveals as if she finds agreement surplus to be worse than a perfect substitute for disagreement surplus.

Subjective inefficiency deals with a harder case. It essentially says that, in some situations, the planner should be willing to give up disagreement surplus in order to increase agreement surplus even if doing so would strictly decrease total surplus (e.g., x_o versus x_2 in Figure 2). If a planner is never willing to do so, then any given reduction in agreement surplus can be compensated by some sufficiently large increase in disagreement surplus. This seems contrary to any normatively reasonable conceptualization of economic efficiency that values mutual consent.[5] Thus, the relevant question is not so much about 'whether' as about 'when': when it may be reasonable to treat disagreement surplus as a 'bad', and not merely as a 'good' that is an imperfect substitute for agreement surplus?

Subjective inefficiency treats disagreement surplus as a bad at a power disparity if it exceeds the agreement surplus at that same power disparity. The size of the agreement surplus at a power

[5] Acemoglu and Wolitzky (2011) extend a principal-agent model by allowing the possibility that the principal can 'coerce' the agent. They demonstrate that if the stronger party must incur a cost to carry out coercion, then utilitarian social welfare is relatively greater under the benchmark of no coercion. The dilemma posed by their results is: Would permitting coercion be considered economically efficient if there was no direct resource cost associated with coercion. Consider a thought experiment. Suppose a planner can *costlessly* assign a fraction of the population to serve as slaves for the rest, and some measure of utilitarian social welfare is maximized when the fraction of slave population is strictly positive. Would this make slavery 'economically' efficient?

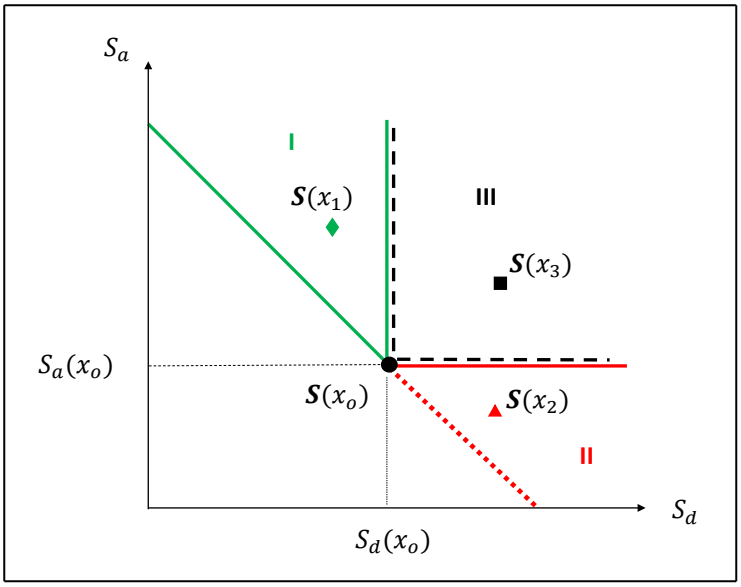

Figure 2: *Illustration of* $\mathbf{D_E}$. Each point denotes a surplus vector with agreement (disagreement) surplus on the vertical (horizontal) axis. The three regions respectively denote that the relatively greater total surplus at some power disparity $y \neq x_o$ can be driven (I) *entirely* by the relatively greater *agreement* surplus at y, or (II) *entirely* by the relative greater *disagreement* surplus at y, or (III) *partly* by both. x_o is objectively inefficient relative to any y whose surplus vector lies in Region I. If $S_d(y) > S_a(y)$ for some y in Region II, then such a y is subjectively inefficient relative to x_o; otherwise, x_o and y are uncomparable. x_o is uncomparable with any y whose surplus vector lies in Region III.

disparity thus serves as the *standard* to determine whether or not disagreement surplus is considered a bad at that power disparity.[6] If disagreement surplus is a bad at a power disparity, then subjective inefficiency recommends giving up disagreement surplus in order to increase agreement surplus even if doing so would decrease total surplus.

2.1.1 Features and implications of $\mathbf{D_E}$

While objective inefficiency seems self-explanatory, it is worth expanding upon the normative idea behind subjective inefficiency. Note that a planner can guarantee total surplus will be at least x by choosing power disparity x. Simple algebra shows disagreement surplus is a bad at x iff the agreement rate at x is strictly below the critical value given by

$$\alpha_b(x) \;=\; \frac{x}{x+\pi}, \;\; \text{such that } \alpha_b(0) = 0, \;\; \alpha_b(\pi) = \frac{1}{2}, \;\; \text{and } \alpha_b'(x) > 0 \;\forall x \in [0, \pi].$$

[6]Subjective inefficiency may be modified by stipulating that disagreement surplus at x is a bad if $S_d(x) > S_a(x) + k(x)$. However, only non-positive values of $k(x)$ seem normatively appealing. Consequently, our choice of $k(x) = 0$ for every x seems conservative. It makes it harder to label a power disparity (subjectively) inefficient relative to any compelling alternative choice of $k(x)$.

A planner can ensure the total surplus will be at least x by choosing power disparity x. The monotonic increase in $\alpha_b(\cdot)$ acts as a hurdle against a planner who is tempted to choose higher levels of power disparity to simply increase total surplus and pays no attention to how it is realized. Since $\alpha_b(x)$ is less than half at each x, disagreements must arise in at least a *strict majority* of all feasible bargaining pairs for disagreement surplus to be considered a bad at any power disparity. Consequently, while subjective inefficiency recommends reducing total surplus if disagreement surplus is considered a bad, a fairly demanding condition must hold for disagreement surplus to be considered a bad in the first place.

The next key feature of $\mathbf{D_E}$ is its silence in two classes of pairwise comparisons. Given a pair of power disparities x and y, neither is inefficient relative to the other as per $\mathbf{D_E}$ if either one of the following two conditions holds.

$[\mathbf{C_1^s}]$ $S_a(x) \geq S_d(x)$, $S_a(y) \geq S_a(x)$, $S_d(y) < S_d(x)$, and $S_t(y) < S_t(x)$.
$[\mathbf{C_2^s}]$ $S_a(y) > S_a(x)$ and $S_d(y) > S_d(x)$.

$\mathbf{C_1^s}$ considers the case where choosing y over x would increase agreement surplus and reduce disagreement agreement surplus and total surplus, but disagreement surplus is not considered a bad at x. It effectively says that there is no compelling justification to choose y over x – i.e., to trade-off disagreement surplus in favor of agreement surplus – *if* disagreement surplus is not considered a bad at x.

$\mathbf{C_2^s}$ refers to such comparisons where greater total surplus is driven partly by agreement surplus and partly by disagreement surplus (e.g., x_o versus any power disparity such as x_3 whose surplus lies in Region III in Figure 2). This silence expresses our inability to answer whether, when, and how much of an increase in the disagreement surplus should be tolerated when it is accompanied with an increase in agreement surplus. It implies an infinitesimally greater agreement surplus at x_3 will render it uncomparable with x_o even if x_3 has arbitrarily greater disagreement surplus than x_o. This feature again highlights that $\mathbf{D_E}$ makes it harder to label a power disparity with greater total surplus inefficient relative to a power disparity with lower total surplus.

Objective and subjective inefficiency induce a transitive, asymmetric, and potentially incomplete ordering of power disparities. They help partition the set of feasible power disparities into efficient and inefficient power disparities.[7] As discussed earlier, $\mathbf{D_E}$ is guided by the idea of 'permissibility', with the (in)efficient power disparities under $\mathbf{D_E}$ being (im)permissible on grounds of economic efficiency. This approach mirrors the judicial scrutiny of economic legislation whereby courts typically uphold an economic legislation as per the *rational basis test* so long as it is a reasonable way to further some "legitimate end ... perfection is by no means required" (Chemerinsky, 2016). $\mathbf{D_E}$ unambiguously treats surplus realized with mutual consent as legitimate and demands strong evidence to consider surplus realized without mutual consent as illegitimate.

[7]This parallels how 'strict Pareto dominance' can be used to order power disparities when utility functions are easily measurable, and end up with a binary categorization of power disparities as Pareto efficient or Pareto inefficient.

Overall, $\mathbf{D_E}$ attempts to make it difficult to label a power disparity inefficient while maintaining the normative distinction between surplus realized with and without mutual consent. This may unduly enlarge the set of power disparities that could be justified on grounds of economic efficiency. The primary implication being that our empirical analysis would be *a priori* biased against refuting the presumption that some efficient power disparities are unfair, regardless of how one defines (un)fair power disparities.

2.2. Defining fair and unfair power disparities

With a slight abuse of notation, let $G(x)$ and $M(x)$ respectively denote the distributions of proposer offers and responder MAOs in $\mathcal{U}(x)$. We refer to an agent whose preferences are defined solely over material payoffs to oneself as *self-regarding*. For expositional ease, any agent who is not self-regarding is called *other-regarding*. A self-regarding proposer will offer nothing to the responder in $\mathcal{U}(\pi)$. Let $G_s(x)$ and $G_o(x)$ respectively denote the distributions of offers in $\mathcal{U}(x)$ by self-regarding and other-regarding proposers. The corresponding expected values of all the four abovementioned distributions will be denoted by $\overline{G}(x)$, $\overline{M}(x)$, $\overline{G_s}(x)$, and $\overline{G_o}(x)$. We define an (un)fair power disparity as follows.

Definition $\mathbf{D_F}$. *Consider a power disparity $x \in [0, \pi]$.*

- *x is dispute-proof if $\overline{G}(x) \geq \overline{M}(x)$.*
- *x is selfish-proof if $\overline{G_s}(x) \geq \overline{G_o}(x)$.*
- *x is unfair if it is neither dispute-proof nor selfish-proof.*
- *x is fair if it is not unfair.*[8]

Rejection by a responder in an ultimatum game is usually interpreted as if she finds the proposer's offer unfair. Dispute-proofness (henceforth, DP) simply transposes this idea to the aggregate level. Violation of DP at x indicates that bargaining between the agents systematically gives rise to disputes because responders find proposers' offers unfair. The MAO of a responder embodies her perspective as to what distributive fairness minimally requires. Hence, DP can be viewed as testing whether the behavior of the proposers (the stronger parties) at a power disparity meets a *standard* of judgement – MAOs of the responders – which embodies the perspective of the weaker parties.

In order to ensure the judgement about unfairness of a power disparity is not based solely on the perspective of the weaker parties, it seems necessary to evaluate the behavior of stronger parties

[8] $\mathbf{D_F}$ is stated as comparing distributions via their expected values for expositional ease. In the empirical analysis, we shall consider multiple tests of difference.

relative to some *other* standard which does not directly rely on the perspective of the weaker parties. There are two broad ways in which one might construct such a standard – a standard derived from the views of uninvolved third-parties, or a standard derived from the behavior of a subset of stronger parties themselves. We follow the latter approach.

If a power disparity is selfish-proof, then on average self-regarding proposers behave no more 'selfishly' than other-regarding proposers. Selfish-proofness (henceforth, SP) is violated at a power disparity when self-regarding proposers systematically offer less than other-regarding proposers. SP may thus be interpreted as evaluating the behavior of self-regarding stronger parties relative to the standard of reasonable conduct provided by the behavior of *other-regarding* stronger parties. Intuitively, when accused of acting unfairly, a common defense is to argue that 'others do the same'. SP is violated at a power disparity precisely when self-regarding stronger parties cannot make this argument since other-regarding stronger parties would be behaving more generously under the same contextual constraints.

Remark 1. *Selfish-proofness as irrelevance of benevolence*: Systematic evidence of benevolence on part of stronger parties will, by definition, be absent at a selfish-proof power disparity. A power disparity becomes too much according to SP when benevolence on part of stronger parties becomes observationally evident and payoff relevant to the weaker parties. Thus, selfish-proofness demands the systematic *irrelevance* of benevolence on part of the stronger parties.[9]

As per $\mathbf{D_F}$, information about the behavior of agents at x suffices to assess whether x is unfair since the two standards for the fairness judgement are derived from agents' behavior at x. The content of the two standards, however, is likely to vary endogenously with changes in power disparity. Average offer by all proposers, average offers by other-regarding and self-regarding proposers, and the average MAO of responders may decline with an increase in power disparity. How the assessment of unfairness of power disparities varies with the level of power disparity will depend on the *relative* rates of decline in offers and MAOs and offers by the two types of proposers.

2.2.1. Motivation behind $\mathbf{D_F}$

DP is inspired primarily by the legal notion of 'reasonable expectations'. In the context of disputes relating to standard-form take-it-or-leave-it contracts, Kessler (1943) noted that "courts have to determine what the weaker contracting party could legitimately expect ... and to what extent the stronger party disappointed the adhering party's reasonable expectations." While there is no

[9]Perfectly competitive markets are often regarded as a benchmark for a fair market structure, presumably because no agent has any market power which eliminates the concern about abuse of power (Landes and Posner, 1978; Trebilcock, 1993; Eisenberg, 1982). Dufwenberg et al. (2011) show that there is no difference in the behavior of self-regarding and other-regarding agents in a perfectly competitive market for a sufficiently rich class of agents' preferences. Selfish-proofness may thus be interpreted as requiring the power disparity to be close enough to the benchmark of a perfectly competitive market so that benevolence on part of stronger parties is indeed irrelevant.

settled definition of what reasonable expectations means, Mitchell (2003) summarizes it as "some entitlement to be treated in a certain way ... an appeal to reasonable expectation is not so much a statement about the actual expectations ... as a judgement of the court ex post facto as to the standards the parties must observe ..." We interpret the MAO of a responder as a behavioral measure of her *reasonable expectation* since she is willing to forego any payoff less than her reported MAO. This ensures we do not have to impose our views as external observers regarding what is the objectively appropriate minimal entitlement of the weaker parties, or what would be objectively reasonable for them to expect.

SP draws upon the notion of 'reasonable conduct' which is routinely invoked across several areas of law (Miller and Perry, 2012). Our formulation of selfish-proofness is inspired by the observation that accounts of reasonable conduct in law – whether trying to highlight its usefulness or its uselessness – seem to treat lack of *indifference* towards others as the minimal content of reasonableness (Herbert, 1935; Moran, 2003). We treat every other-regarding proposer – one who offer a strictly positive amount in $\mathcal{U}(\pi)$ and reveals he is not indifferent towards the responder – as a 'reasonable' proposer. Put differently, a reasonable proposer is one who does not completely exploit his bargaining power to his own material advantage when his power is at its peak. Our choice to utilize the behavior of other-regarding proposers to construct the standard of reasonable conduct is broadly consistent with the view that such standards combine "judgments about how we should behave with determinate facts about how we do behave ... focusing our attention on some actual practice that has won our approval as an appropriate standard to guide our conduct" (Scalet, 2003).

2.2.2. Alternatives to $\mathbf{D_F}$

While $\mathbf{D_F}$ considers it necessary to account for the perspective of both stronger and weaker parties, even the simultaneous violation of both DP and SP may be considered insufficient to label a power disparity unfair. For instance, in appealing to fairness contract law seeks to avoid 'egregiously' unfair bargains. One example is the doctrine of unconscionability that permits judicial intervention in private contracts when the contractual terms are so unfair as to "shock the conscience of the court" (see, for e.g. Farnsworth, 1982 Chapter 4.E). Hence, one way to formulate an alternative definition of an unfair power disparity would be to supplement $\mathbf{D_F}$ with other conditions that must hold in addition to the violation of both DP and SP.

If one believes violation of DP and SP should be considered insufficient to label a power disparity unfair, then $\mathbf{D_F}$ would unduly label some fair power disparities unfair. As this would potentially enlarge the set of unfair power disparities, it will bias our empirical analysis towards supporting the presumption that some efficient power disparities are unfair. Consequently, we believe that for the purposes of the present paper, the main weakness of $\mathbf{D_F}$ lies in that it does not pin-point a compelling way to identify the 'reasonable' agents whose behavior informs the standard of reasonable conduct. Hence, in assessing the robustness of our findings, we shall consider the alternatives to $\mathbf{D_F}$ that utilize alternative identification strategies. More importantly, we shall focus on such

identification strategies that make the standards more demanding. They are expected to make it easier to label a power disparity unfair, and thus harder to refute the presumption that some efficient power disparities are unfair.

2.3. Patterns of inefficient versus unfair power disparities

Whether a power disparity x is fair or unfair depends on the behavior of agents at only x. In contrast, whether a power disparity is efficient or inefficient may depend on the behavior of agents at all power disparities. A model where it is common knowledge that all agents are self-regarding will predict (a) no disagreements at any power disparity and (b) no violation of dispute-proofness or selfish-proofness at any power disparity $x \in [0, \pi]$. Consequently, under the assumption of self-regarding preferences every power disparity is efficient as per $\mathbf{D_E}$ and fair as per $\mathbf{D_F}$.

Now consider a model where a fraction of agents are self-regarding and the remainder are inequity-averse (Fehr and Schmidt, 1999). Average MAO of responders will converge to zero as x approaches π since the ability of a responder to materially punish the proposer by rejecting an offer declines with an increase in x and completely vanishes at $x = \pi$. This guarantees there can be no disagreements in $\mathcal{U}(\pi)$, thereby rendering $x = \pi$ efficient as per $\mathbf{D_E}$. In addition, with a strictly positive fraction of other-regarding types, average proposer offer in $\mathcal{U}(\pi)$ will converge to some $\bar{G}(\pi) > 0$. Consequently, there exists a threshold power disparity $\hat{x} < \pi$ such that every power disparity beyond $\hat{x}$ will be dispute-proof; and hence, fair as per $\mathbf{D_F}$. Thus, such a model predicts (a) the highest power disparity, $x = \pi$, is efficient and (b) sufficiently high power disparities are fair. A planner who makes the abovementioned preference assumptions will not hesitate in choosing the highest level of power disparity on grounds of economic efficiency as per $\mathbf{D_E}$, or on the grounds of fairness as per $\mathbf{D_F}$.

Notice that the above predictions hinge on average responder MAO converging to zero as x approaches π. While the ability of a responder to materially punish the proposer declines with an increase in x, a responder may nonetheless report MAOs that are bounded away from zero even at sufficiently high levels of x for *expressive* reasons. There is no dearth of studies providing a rationale for and evidence of expressive behavior (see Hillman 2010). In light of this, consider the following assumption.

[$\mathbf{A}^*$] *Agents' preferences are such that the following patterns of behavior are predicted for* $\mathcal{U}(x)$.

(a) $\overline{G}(x)$, $\overline{G_o}(x)$, $\overline{G_s}(x)$, *and* $\overline{M}(x)$ *decrease with an increase in* x.

(b) $\overline{G}(0) \geq \overline{M}(0)$.

(c) $\overline{G_o}(0) \geq \overline{G_s}(0)$.

(d) $\overline{M}(\pi) > 0$.

The first three parts contain minimal assumptions consistent with existing empirical evidence (Kagel and Cooper, 2016) and the last part reflects the assumption that some agents have expressive preferences. In the following we describe that under $\mathbf{A}^*$,

- no sufficiently general predictions can be made regarding which power disparities will be (in)efficient; but,
- sufficiently low power disparities are fair whereas sufficiently high power disparities can be fair or unfair.

2.3.1. Inefficient power disparities

We first describe that whether a particular power disparity is deemed efficient or inefficient under $\mathbf{D_E}$ potentially depends on the agreement rates at *all* power disparities. Then we use a simple example consistent with $\mathbf{A}^*$ to illustrate that the richness of feasible patterns of (in)efficient power disparities precludes any sufficiently general conclusions about which power disparities will be (in)efficient.

Suppose the agreement rate at a power disparity $x_o \in [0, \pi]$ is $\alpha_o \in [0, 1]$. Agreement surplus at a power disparity y will be weakly greater than the agreement surplus at x_o if $\alpha(y)\pi \geq \alpha_o\pi$, which requires $\alpha(y) \geq \alpha_o$. Disagreement surplus at y will be weakly lower than the disagreement surplus at x_o if $(1 - \alpha(y))y \leq (1 - \alpha_o)x_o$, which requires

$$\alpha(y) \;\geq\; \alpha_d(y|x_o, \alpha_o) \;=\; \alpha_o + (1 - \alpha_o)\Big(1 - \frac{x_o}{y}\Big).$$

Similarly, total surplus will be weakly greater at y than at x_o if

$$\alpha(y) \;\geq\; \alpha_t(y|x_o, \alpha_o) \;=\; \alpha_o + (1 - \alpha_o)\Big(\frac{x_o - y}{\pi - y}\Big).$$

Straightforward algebra suggests x_o will be objectively inefficient relative to y if

$$\alpha(y) \;\geq\; \begin{cases} \alpha_t(y|x_o, \alpha_o) & \text{for } y < x_o. \\ \alpha_d(y|x_o, \alpha_o) & \text{for } y > x_o. \end{cases} \tag{0.1}$$

Similarly, x_o will be subjectively inefficient relative to y if $\alpha_o < \alpha_b(x_o)$ and $\alpha(y) \in \big[\alpha_o,\ \alpha_t(y|x_o, \alpha_o)\big)$.

Note that a power disparity x_o can be objectively inefficient relative to a $y < x_o$ or $y > x_o$ if the agreement rate at y is sufficiently greater than the agreement rate at x_o. In contrast, a power disparity x_o can be subjectively inefficient relative to a $y < x_o$ that has relatively greater agreement rate; but, it cannot be subjectively inefficient relative to any $y > x_o$ regardless of the pattern of variation in agreement rates. This is because no $y > x_o$ can simultaneously have greater agreement surplus and lower total surplus than x_o.

Example 1. *Suppose the aggregate agreement rate at $x \in [0,\pi]$ is given by $\alpha(x) = \alpha_o - \beta\frac{x}{\pi}$, where $\beta \in [0,\alpha_o]$ for any $\alpha_o \in [0,1]$. Let x_t denote the power disparity such that $S_t(0) = S_t(x_t)$ and x_{ad} be such that $S_a(x_{ad}) = S_d(x_{ad})$. The set of inefficient power disparities is*

$$X_{IE} = \begin{cases} \emptyset & \text{if } S_a(\pi) \geq S_d(\pi) \text{ and } S_t'(0) \geq 0 \\ (x_{ad},\pi] & \text{if } S_a(\pi) < S_d(\pi) \text{ and } S_t'(0) \geq 0 \\ (0,x_t] & \text{if } S_a(\pi) \geq S_d(\pi) \text{ and } S_t'(0) < 0 \\ (0,x_t] \bigcup (x_{ad},\pi] & \text{if } S_a(\pi) < S_d(\pi),\ S_t'(0) < 0, \text{ and } x_t < x_{ad} \\ (0,\pi] & \text{if } S_a(\pi) < S_d(\pi),\ S_t'(0) < 0, \text{ and } x_t \geq x_{ad} \end{cases} \tag{0.2}$$

α_o is the agreement rate in $\mathcal{U}(0)$ and β captures the rate of decline in agreement rates.[10] Figure 3 illustrates the subset of power disparities that are deemed inefficient under $\mathbf{D_E}$ depending upon these two parameters. Note that virtually all or no power disparity may be inefficient (Region I versus Region V in Panel A); the set of inefficient power disparities may be disconnected or connected (Region IV versus any other region); and, inefficiency may be concentrated at high or low power disparities (Region II versus Region III). For any (α_o,β), which power disparities in $[0,\pi]$ turn out to be inefficient is broadly determined by a combination of two factors. If rate of decline in agreement rates is sufficiently large to ensure $S_a(\pi) < S_d(\pi)$, then some sufficiently high power disparities will necessarily be subjectively inefficient relative to $x = 0$. If the rate of decline in the agreement rates is sufficiently large to ensure $S_t'(0) < 0$, then some sufficiently low strictly positive power disparities will necessarily be objectively inefficient relative to $x = 0$. This example highlights that $\mathbf{D_E}$ leaves substantial room for agents' behavior to determine which power disparities are deemed (in)efficient.

2.3.2. Unfair power disparities

In general, if $\overline{G}(0) \geq \overline{M}(0)$, then sufficiently low power disparities are *likely* to be dispute-proof. However, sufficiently high power disparities may or may not be dispute-proof since average responder MAO may decline relatively faster than average proposer offer with an increase in power disparity. Theoretical considerations cannot resolve whether $\overline{G}(\pi)$ is greater or lower than $\overline{M}(\pi)$. This is ultimately an empirical question which depends on the prevalence and intensity of expressive preferences among responders and other-regarding preferences among proposers.

A relatively lower power disparity acts like a relatively stronger constraint on the proposers as it makes *both* types of proposers relatively more fearful of rejection by responders. As power disparity increases, this constraint weakens and it becomes relatively more likely that a proposer's preference-type gets revealed in his offer. Hence, sufficiently high power disparities are *unlikely* to

[10] We use this example because results from some existing studies suggest agreements rates are likely to decline with an increase in power disparity (Hennig-Schmidt et al., 2018).

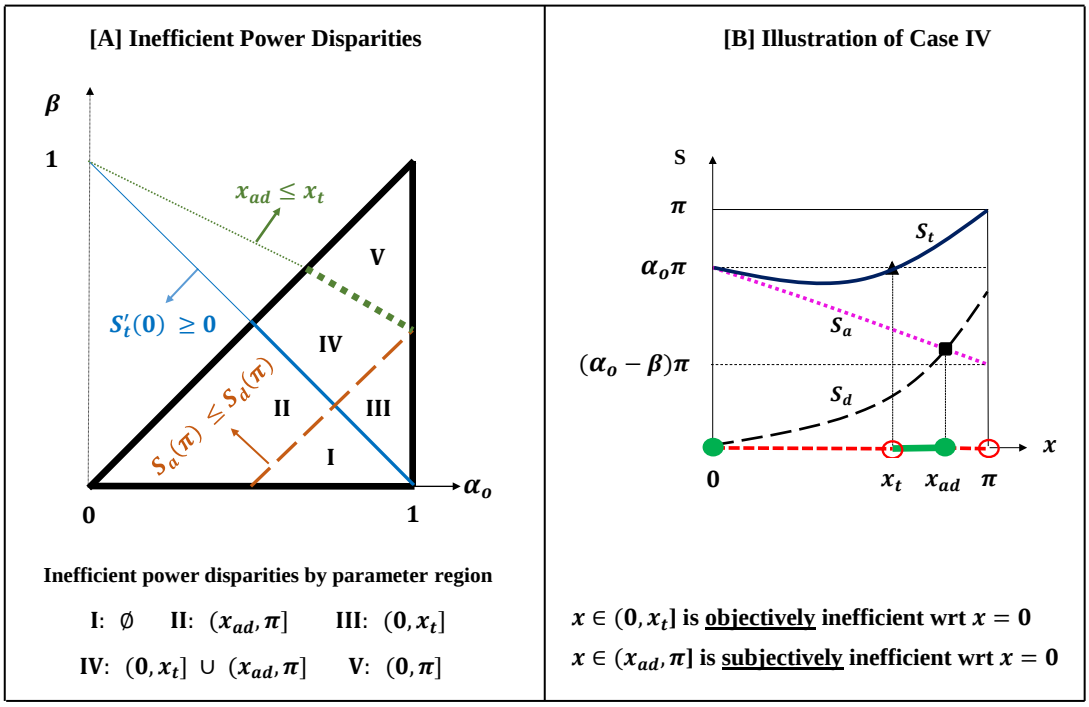

Figure 3: *Patterns of inefficient power disparities for linearly declining agreement rates (Example 1).* Panel A illustrates the set of inefficient power disparities at each feasible combination of (α_o, β). The combinations of (α_o, β) that are feasible given declining agreement rates lie in or on the triangle formed by the thick lines. Panel B illustrates the details for an (α_o, β) in Region IV. The level of power disparity is shown on the horizontal axis and the three surplus values – total, agreement, and disagreement – appear on the vertical axis. Efficient power disparities are indicated with thick lines or closed circles. Inefficient power disparities are indicated with dashed lines or open circles. Each $x \in (0, x_t]$ is objectively inefficient wrt every $x = 0$. Each $x \in (x_{ad}, \pi]$ is subjectively inefficient wrt every $y \in [0, x_{ad}]$. No $x \in (x_t, x_{ad}] \bigcup \{0\}$ is subjectively or objectively inefficient. The set of inefficient power disparities is non-empty and disconnected.

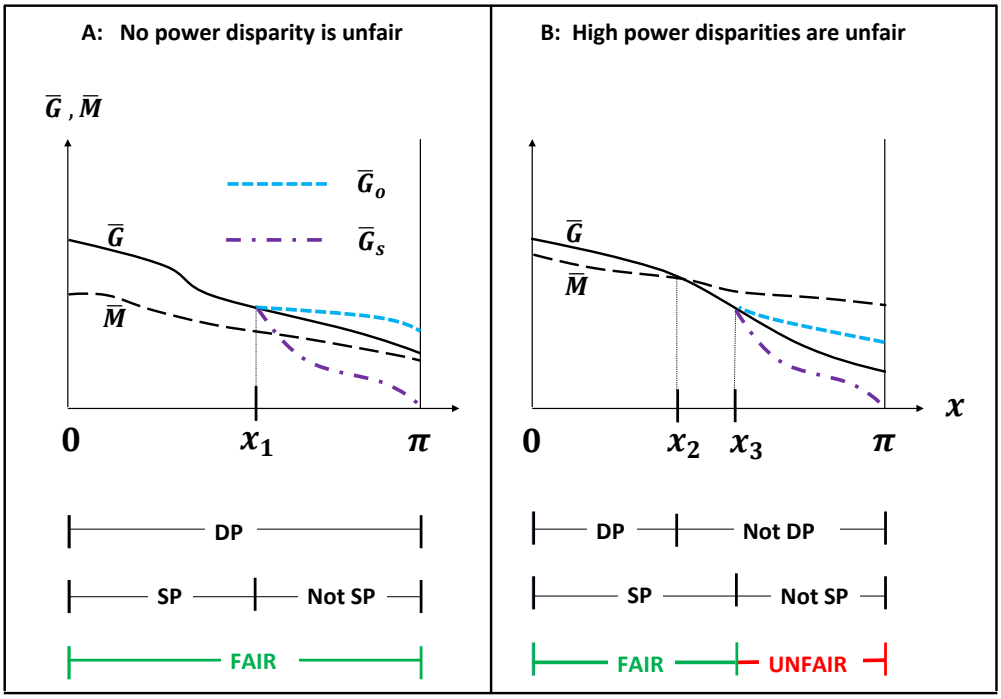

Figure 4: *Illustration of D_F.* Average offers and MAOs illustrated in both Panels are consistent with $\mathbf{A}^*$. In Panel A, $\bar{G}(x) \geq \bar{M}(x)$ holds at every $x \in [0, \pi]$ and thus every power disparity is dispute-proof. $\bar{G}_s(x) \geq \bar{G}_o(x)$ holds only for $x \in [0, x_1]$ and thus power disparities in $(x_1, \pi]$ are not selfish-proof. Every power disparity is fair since *simultaneous* violation of DP and SP does not occur at any $x \in [0, \pi]$. In Panel B, every $x \in (x_3, \pi]$ is unfair since both DP and SP are violated. No $x \in [0, x_3]$ is unfair as DP or SP holds.

be selfish-proof. By definition, $x = \pi$ cannot be selfish-proof. Sufficiently low power disparities, however, may or may not be selfish-proof. For instance, it remains an open question whether other-regarding proposers offer on average *strictly more* or *no more* than self-regarding proposers in the standard ultimatum game $\mathcal{U}(0)$ (Thaler, 2016 pp. 142).

The above discussion suggests that under $\mathbf{A}^*$ sufficiently low power disparities are *likely* to be dispute-proof and sufficiently high power disparities are *unlikely* to be selfish-proof. A power disparity is unfair under our definition if it is neither dispute-proof nor selfish-proof. Theoretical models where $\mathbf{A}^*$ holds will therefore predict sufficiently low power disparities will be fair because DP and SP will not be simultaneously violated. Sufficiently high power disparities may however be predicted to be fair or unfair depending upon the specific assumptions made by the analyst (see Figure 4 which illustrates two simple cases where the behavioral patterns are consistent with $\mathbf{A}^*$ and with declining agreement rates). These predictions, in conjunction with the ambiguity in predicting which power disparities will be (in)efficient, explain why it is difficult to sharply predict whether and when the tension between efficiency and fairness in the evaluation of power disparities will actually arise.

3. Experimental Implementation

The experiment involves two treatments. Treatment $\mathcal{U}$ is based on $\mathcal{U}(x)$. In treatment $\mathcal{CU}$, the first mover can *choose* either to take an outside option or to enter $\mathcal{U}(x)$ and bargain over a bigger under the knowledge that she will be the weaker bargaining party (i.e., the responder). The experiment involved 12 sessions which were conducted at the Monash Laboratory for Experimental Economics in the Department of Economics between October 2014 and March 2015. Subjects were recruited through ORSEE (Greiner, 2004). The experiment was programmed in *zTree* (Fischbacher, 2007). Sessions lasted no more than 50 minutes and average payment received by subjects was AUD 21. A total of 230 students from various faculties participated (120 in treatment $\mathcal{U}$ and 110 in treatment $\mathcal{CU}$). In the following, we describe the details of the two treatments and our empirical strategy to assess whether any efficient power disparity is unfair.

3.1. Treatment $\mathcal{U}$

Treatment $\mathcal{U}$ was implemented for six values of $x \in X = \{0, 12, 18, 23, 27, 30\}$ using a pie size of $\pi = 30$. A within-subject design was utilized so that we can identify self-regarding and other-regarding proposers using their offers in $\mathcal{U}(30)$ which, in turn, helps assess whether power disparities other than the highest power disparity, $x = 30$, are selfish-proof.[11]

Treatment $\mathcal{U}$ involves two blocks – the ultimatum block with six interactions that correspond to $\mathcal{U}(x)$ for the six different values of values of $x \in X$; and, the dictator block that involves one dictator game with a pie size of $\pi = 30$. The dictator game, henceforth $\mathcal{D}$, is included to provide an additional way to identify self-regarding and other-regarding proposers. The order of the ultimatum and dictator blocks is randomized across sessions.[12] The instructions informed subjects about both the blocks at the start of the experiment. Consider a session where subjects first play the ultimatum block. After instructions and the comprehension quiz, the sequence of events in such a session is as follows.

[1] Role Assignment for $\mathcal{U}$ and $\mathcal{D}$. Subject are randomly assigned to one of two roles: either the proposer for the six ultimatum games and dictator for the dictator game, or the responder for the six ultimatum games and the recipient for the dictator game.

[2.1] Elicitation of offers and MAOs in $\mathcal{U}$. Proposers report their offers in $\mathcal{U}(x)$ for each $x \in X$ on a single decision screen. Similarly, responders report their MAOs in $\mathcal{U}(x)$ for each $x \in X$ on a

[11] The experimental instructions and screenshots are provided in Appendix 3. The instructions were framed in a neutral language. Subjects answered a comprehension quiz before starting the experiment. The chosen values of x are disproportionately concentrated towards higher values because we anticipated the tension between fairness and efficiency would be unlikely to arise at sufficiently low levels of power disparity.

[12] Fligner-Policello robust rank-order tests reveal the order in which the two blocks were implemented has no significant effect on offers or MAOs (Table 6 in Appendix 2). Hence, we pool the data while reporting the results.

single decision screen. No subject receives any information about the decisions by other subjects while making decisions.

[2.2] Belief Elicitation for $\mathcal{U}$. For each $x \in X$, each proposer is asked to guess the number of responders in the session whose MAOs are no more than his offer. This provides a proposer's perceived likelihoods that his actual offers at different power disparities would be accepted. Similarly, each responder is asked to guess the number of proposers in the session whose offers are no less than her MAO. This provides a responder's perceived likelihoods that her MAOs at different power disparities would be satisfied. Subjects were paid for the accuracy of their beliefs but they were not informed about the belief elicitation in the instructions. For each guess, a subject earns AUD 1 if the guess is within 1 unit of the correct answer.

[3] Giving in $\mathcal{D}$. Subjects in the role of dictator decide how much to give to a recipient out of π. As mentioned earlier, the primary purpose of including the dictator game is to have an additional way to identify self-regarding or other-regarding proposers. Our preferred identification relies of offers by proposers in $\mathcal{U}(30)$ which is arguably 'closer' to the environment of interest. The absence of any choice on part of the recipient in the dictator game is believed to introduce considerations that are absent in ultimatum games.[13] Hence, we shall utilize the identification of proposers based on their dictator game giving to assess the robustness of our findings.

[4] Demographic questionnaire and payment. Once all stages are over, a subject is randomly invited to draw one card from a pack of seven labeled Stage 1 to Stage 7. All subjects get paid in accordance with their decisions in the drawn stage. Subjects additionally receive a AUD 5 show-up fee and their earnings from the belief elicitation stage. Subjects are requested to fill a demographic survey prior to receiving payment. Finally, subjects are paid individually by an experimenter and discharged.

3.2. Treatment $\mathcal{CU}$

Treatment $\mathcal{U}$ considers the ex-post bargain between two agents *in medias res* i.e., already locked in an exchange. If agents find themselves in such an ex-post bargain, it is likely they initiated their transaction with mutual consent at some point in the past. Explicitly accounting for the voluntary choice to initiate the transaction may affect the likelihood of an ex-post dispute as it may influence the expectations of the weaker party and the beliefs of the stronger party regarding what the weaker party may find reasonable to expect.

[13]The dictator game is contextually distinct from the ultimatum game since the recipient cannot influence the outcome. van Dijk and Vermunt (2000) and Handgraaf et al. (2008) find that this feature cues the dictator to be more concerned about the recipient. Note that even in $\mathcal{U}(30)$ where the power disparity is highest, the responder can influence the outcome.

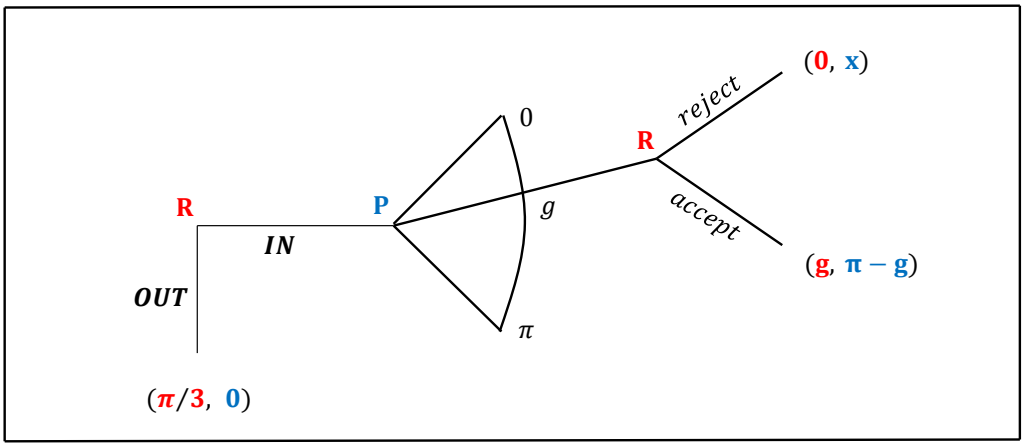

Figure 5: $\mathcal{CU}(x)$, *where* $x \in [0, \pi]$. The first mover chooses whether to enter the bargain $\mathcal{U}(x)$ and occupy the role of the (R)esponder, or to take an outside option and not initiate the interaction with (P)roposer. The first (second) payoff entry at each terminal node corresponds to the payoff of R (P).

Treatment $\mathcal{CU}$ supplements treatment $\mathcal{U}$ by allowing the ex-post weaker party to ex-ante choose whether or not to initiate the transaction. Consider the game, $\mathcal{CU}(x)$, where the first mover can choose (a) either to enter the bargain $\mathcal{U}(x)$ and occupy the role of the responder or (b) to take an outside option and not initiate the transaction. Since the first mover occupies the role of the responder upon choosing to enter the $\mathcal{U}(x)$ subgame, with some abuse of the label, we shall henceforth refer to the first mover as the responder (Figure 5).

If the responder goes OUT, the game ends with a payoff of zero for the proposer and AUD 10 for the responder. If the responder goes IN, the game proceeds to the ex-post bargaining stage as per $\mathcal{U}(x)$. $\mathcal{CU}(x)$ combines the key features of the trust game (Berg et al., 1995) and the modified ultimatum game $\mathcal{U}(x)$. The experimental implementation of treatment $\mathcal{CU}$ closely follows that of treatment $\mathcal{U}$. For each of the six values of $x \in X$, responders choose IN or OUT. A responder reports her MAOs for only those values of x at which she chooses IN. Proposers, however, are neither informed about the entry choices of responders nor their MAOs conditional on entry, and asked to report their offer for each $x \in X$.

Subjects in treatment $\mathcal{CU}$ also play a standard trust game, with responders as trustors and proposers as trustees. Each trustor chooses either to take a payoff of AUD 10 by going OUT or to go IN. Trustees choose the amount to return to the trustor out of $\pi =$ AUD 30, without knowing the entry choice of the trustor. Thus, treatment $\mathcal{CU}$ also contains a total of seven decision stages. The standard trust game is played before or after the block containing six $\mathcal{CU}(x)$ games. One out of the seven stages is randomly chosen for payment.[14]

[14]The structure of $\mathcal{CU}$ prevents eliciting the same type of beliefs as in $\mathcal{U}$. Proposers guess whether the matched responder chose IN or OUT, and earn AUD 1 for a correct guess. Responders guess the offer by the matched responder at every $x \in X$ regardless of their entry choice, and earn AUD 1 for each guess that is within 1 unit of the actual offer.

We assess whether a power disparity x is unfair or inefficient in treatment $\mathcal{CU}$ by restricting attention to only those responders who choose to enter.[15] This ensures our inferences are based on the *actual exercise* of power by the proposers in treatment $\mathcal{CU}$. One would expect entry rates to decline with an increase in x. We will be unable to make efficiency or fairness assessments of an $x \in X$ if all responders choose OUT. Differences between treatments $\mathcal{CU}$ and $\mathcal{U}$ are likely to be driven by greater expectations of the weaker parties due to the certain payoff foregone upon entry. Whether proposers meet their increased expectations is an empirical question. Thus, whether and when the tension between efficiency and fairness arises in treatment $\mathcal{CU}$, and how it differs relative to treatment $\mathcal{U}$, remains theoretically ambiguous.

3.3. Empirical strategy

Our empirical analysis will focus on investigating the presumption that some efficient power disparities can be unfair. The tests of objective and subjective inefficiency involve multiple comparisons of surplus values. We operationalize these tests by first utilizing the offer by any proposer $p \in P$ and MAOs of all responders at an $x \in X$ to construct the agreement rate for a proposer $p \in P$ at $x \in X$ as described in Section 2. The agreement surplus, disagreement surplus and total surplus with respect to proposer p can then be calculated.[16] The distributions of these three surplus values at each $x \in X$ are constructed by repeating this procedure for every proposer.

As per definition $\mathbf{D_E}$, the conditions for establishing that power disparity x is (subjectively or objectively) inefficient relative to power disparity y involve checking for a combination of strict and weak inequalities. Hence, the evidence for x being inefficient relative to y requires refuting a combination of strict and weak inequalities. For consistency in reporting the results of our statistical tests, we set the null hypothesis as the two distributions being drawn from the same population and report two-tailed p-values for the Wilcoxon Signed-Rank test.

We shall follow the same strategy while testing for violations of dispute-proofness or selfish-proofness at any $x \in X$ as per definition $\mathbf{D_F}$. Here we use the Fligner-Policello robust rank-order test and the Kolmogorov-Smirnov test as choices across different sets of subjects have to be compared. We shall also assess violations of DP and SP by testing for differences in the relevant means via regression analysis. As a violation of DP requires average responder MAO to be greater than average proposer offer, we test for violations of DP using

$$Y_{ix} \;=\; \alpha \;+\; \sum_{x \in X} \beta_x \mathbb{I}_x \;+\; \sum_{x \in X} \gamma_x \big(\mathbb{I}_x \cdot \mathbb{I}_i^p\big) \;+\; \delta \mathbf{Z}_i \;+\; \epsilon_i \tag{0.3}$$

[15] Incorporating the non-entering responders in the efficiency assessments – to account for the impact of the *threat* that the stronger party can exploit his power advantage – is feasible. However, how to incorporate the non-entering responders into the fairness assessments is unclear given the lack of information about their MAOs.

[16] Recall, for any proposer $p \in P$ with agreement rate $\alpha^p(x)$ at $x \in X$, the agreement, disagreement and total surplus are respectively $S_a^p(x) = \alpha^p(x)\pi$, $S_d^p(x) = (1 - \alpha^p(x))x$, and $S_t^p(x) = S_a^p(x) \;+\; S_d^p(x)$.

where i ranges over all the subjects; Y_{ix} equals the offer (MAO) by subject i at $x \in X$ if subject i is a proposer (responder); the binary indicator $\mathbb{I}_x$ takes the value 1 iff the observation involves a decision made at power disparity $x \in X$; the binary indicator $\mathbb{I}_i^p$ takes the value 1 iff subject i is a proposer; and, $\mathbf{Z}$ is a vector of demographic characteristics of subjects. The coefficients of interest are $\{\gamma_x\}_{x \in X}$ which give the average difference between proposer offers and responder MAOs at each power disparity, with a negative estimate of γ_x indicating violation of dispute-proofness at $x \in X$.

As a violation of SP requires the average offer by self-regarding proposers to be lower than the average offer by other-regarding proposers, we test for violations of SP using

$$g_{px} \;=\; \alpha \;+\; \sum_{x \in X} \beta_x \mathbb{I}_x \;+\; \sum_{x \in X} \eta_x \big(\mathbb{I}_x \cdot \mathbb{I}_p^{SRP}\big) \;+\; \delta \mathbf{Z}_p \;+\; \epsilon_p \tag{0.4}$$

where p ranges over all subjects who act as a proposer; g_{px} is the offer by proposer $p \in P$ at $x \in X$; the binary indicator $\mathbb{I}_p^{SRP}$ takes the value 1 iff proposer p is identified as a self-regarding proposer on the basis of his offer at $x = \pi$ being zero. The coefficients of interest are $\{\eta_x\}_{x \in X}$ which give the average difference in offers by self-regarding and other-regarding proposers at each power disparity, with a negative estimate of η_x indicating violation of selfish-proofness at $x \in X$.

Both specifications are estimated using a random effects regression with standard errors clustered around subjects. The test of DP for treatment $\mathcal{U}$ is based on data from an equal numbers of proposers and responders. In contrast, the number of actual responders at different values of x in treatment $\mathcal{CU}$ depends on the entry choice of the first movers. The number of proposers however remains constant as they were asked to report offers at every $x \in X$. As with the non-parametric tests, we shall test the null hypothesis of γ_x or η_x being zero and report two-tailed p-values.

4. Results

The assessment of power disparities from the fairness and efficiency perspectives may exhibit a variety of qualitative patterns with none, some, or all efficient power disparities being unfair. In the following, we show that the assessment of power disparities from both perspectives exhibits a qualitatively similar *threshold* pattern. In addition, we find no efficient power disparity is unfair in treatment $\mathcal{U}$. These findings continue to hold in treatment $\mathcal{CU}$ where the first mover can choose whether or not to enter a bargain where she will be the weaker bargaining party. Finally, we establish the robustness of these findings.

4.1. Treatment $\mathcal{U}$

Figure 6 depicts the mean surplus values in treatment $\mathcal{U}$. As power disparity increases, mean agreement surplus declines, but mean disagreement surplus and mean total surplus increase. These means suggest that no power disparity is objectively inefficient in our data since a relatively higher total surplus is always accompanied with a relatively lower agreement surplus. For a power disparity

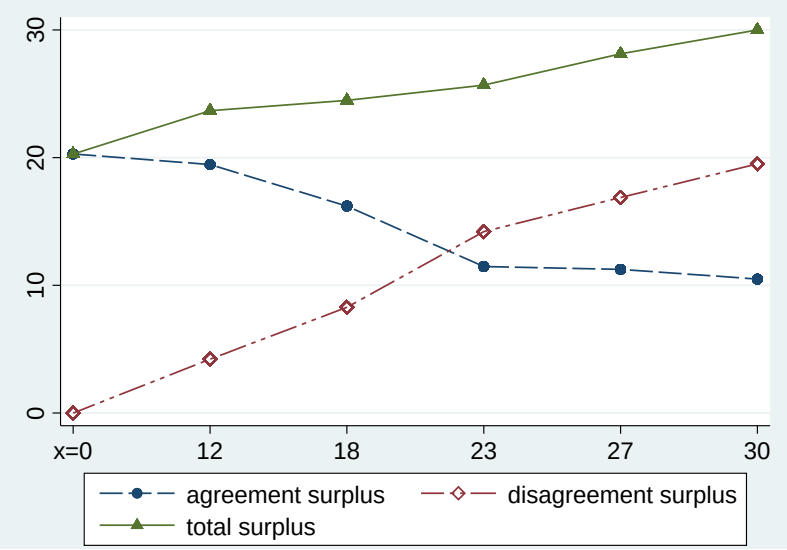

Figure 6: *Mean surplus in treatment* $\mathcal{U}$. Mean surplus comparisons suggest no power disparity is objectively inefficient since for any pair of power disparities a relatively higher total surplus is always accompanied with a relatively lower agreement surplus. The mean disagreement surplus exceeds the mean agreement surplus at every $x \geq 23$; and, the mean disagreement surplus is lower than the mean agreement surplus at every $x \leq 18$. Hence, every $x \geq 23$ is subjectively inefficient relative to each $x \leq 18$ according to mean comparisons.

to be subjectively inefficient disagreement surplus at that power disparity must be a bad, which requires disagreement surplus to exceed the agreement surplus. Figure 6 indicates disagreement surplus is a bad at every $x \geq 23$ according to mean comparisons. Given that mean agreement surplus monotonically decreases but mean disagreement and mean total surplus monotonically increase with x, every $x \geq 23$ is subjectively inefficient.[17] Furthermore, mean surplus comparisons suggest that efficient and inefficient power disparities constitute disjoint connected subsets of the set of feasible power disparities.

We conduct Wilcoxon Signed Rank tests of differences to verify the statistical validity of the above findings. The p-values for the test of whether disagreement surplus is a bad at a power disparity are reported in Table 1. The patterns inferred via mean comparisons hold statistically, with one exception: we cannot reject the hypothesis that total surplus at $x = 12$ and 18 are equal (see Table 2 in Appendix 2 for details). Hence, the possibility remains that $x = 18$ is objectively inefficient relative to $x = 12$. Since we consider a power disparity efficient unless it can be proven inefficient, the above findings can be summarized as follows.

Result [1] *In treatment* $\mathcal{U}$, $x \in \{0, 12, 18\}$ *is efficient and* $x \in \{23, 27, 30\}$ *is inefficient.*

[17] For example, consider the comparison between $x = 0$, and $x = 30$ where disagreement surplus is strictly greater than agreement surplus. Agreement surplus is relatively lower at $x = 30$ by roughly 10 units. But, disagreement surplus is relatively greater by roughly 20 units. Hence, the entire increase in total surplus at $x = 30$ relative to $x = 0$ is driven by its relatively greater disagreement surplus.

Table 1: Tests of difference for treatment $\mathcal{U}$

Null hypothesis	**Power disparity** x					
	0	12	18	23	27	30
NBAD: *S_d is not a bad at x*						
Sign Rank test	–	0.01	0.01	**0.07**	**0.02**	**0.01**
DP: *x is dispute proof*						
Regression test	0.06	0.28	0.83	**0.01**	**0.01**	**0.01**
Fligner-Policello test	0.05	0.25	0.68	**0.01**	**0.01**	**0.01**
Kolmogorov-Smirnov test	0.08	0.66	0.93	**0.03**	**0.02**	**0.01**
SP: *x is selfish proof*						
Regression test	0.37	0.36	**0.01**	**0.01**	**0.01**	–
Fligner-Policello test	0.41	0.73	**0.05**	**0.01**	**0.01**	–
Kolmogorov-Smirnov test	0.58	0.98	0.35	**0.03**	**0.01**	–

Notes. The table reports *two-tailed* p-values for the null hypothesis of equality between the two relevant variables for each test. NBAD is based on the distributions of agreement and disagreement surplus across proposers (n = 60). DP is based on the distributions of offers by the 60 proposers and MAOs of the 60 responders. SP is based on the distributions of offers by the 27 other-regarding and the 33 self-regarding proposers. A proposer with a strictly positive (zero) offer in $\mathcal{U}(30)$ is classified other-regarding (self-regarding). Entries are indicated in bold when the difference is significant at the 10% level *and* the sign of the difference (test statistic) is in the direction required by the corresponding definition. p-values strictly less than 0.01 are rounded off to 0.01.

We now turn to our key question: Is any efficient power disparity unfair? This requires some efficient $x \in \{0, 12, 18\}$ be neither dispute-proof nor selfish-proof. Dispute-proofness asks whether offers by proposers meet the standard of reasonable expectations determined by the MAOs of responders. Figure 7 depicts that both the mean offer by proposers and the mean MAO of responders tend to decline with an increase in power disparity, with offers declining faster than MAOs. The mean MAO is higher than the mean offer at every $x \geq 23$. Statistical tests confirm that dispute-proofness can be rejected at every $x \geq 23$ (Table 1 reports the p-values). Table 8 in Appendix 2 provides the detailed results.

Selfish-proofness tests whether offers by self-regarding proposers meet the standard of reasonable conduct determined by the offers of other-regarding proposers. We first identify self-regarding and other-regarding proposers depending upon whether a proposer offers zero or a strictly positive amount in $\mathcal{U}(30)$. Figure 8 indicates that both mean offers by the 27 other-regarding proposers and mean offers by the 33 self-regarding proposers decline as power disparity increases. At low power disparities, proposers' fear of rejection by responders is high enough to ensure no statistical difference in offers between the two types of proposers. As power disparity increases, this fear of rejection declines and we can reject selfish-proofness at every $x \geq 23$ (see Table 1 for p-values, and Table 4 in Appendix 2 for details).

One caveat is that two out of the three tests suggest $x = 18$ is not selfish-proof. However, since our definition requires a violation of both dispute-proofness and selfish-proofness for a power

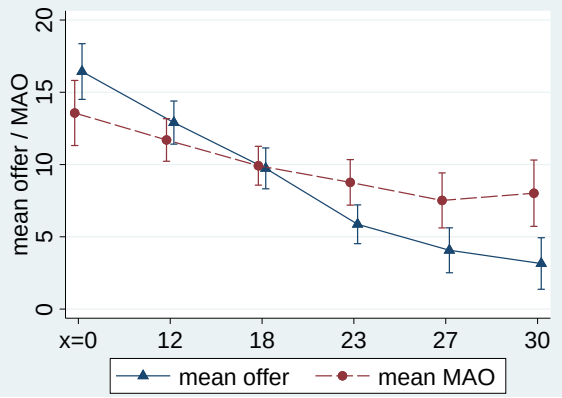

Figure 7: *Dispute-proofness in treatment $\mathcal{U}$.* Mean proposer offer declines with an increase in power disparity. Mean responder MAO largely declines but at a relatively lower rate.

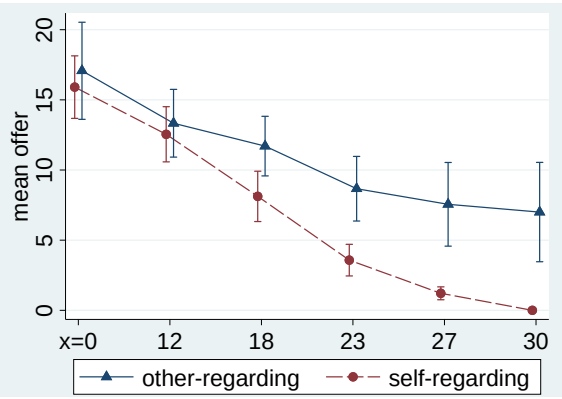

Figure 8: *Selfish-proofness in treatment $\mathcal{U}$.* As power disparity increases, mean offers by both types of proposers decline but the rate of decline is relatively lower for other-regarding proposers.

disparity to be considered unfair, every $x \leq 18$ is fair and every $x \geq 23$ is unfair. Taken together, the above patterns in the assessment of efficient and fair power disparities lead to our central result.

Result [2] *No efficient power disparity in X is unfair in treatment $\mathcal{U}$.*

A key driver of these patterns is that responder MAOs do not converge to zero as x approaches π. The majority of responders (35 out of 60) report a strictly positive MAO in $\mathcal{U}(30)$. These responders may be viewed as having expressive preferences since they are willing to reject a strictly positive offer while knowing that their rejection will *not* materially hurt the proposer. While both average offers and average MAOs largely decline with an increase in power disparity, average MAOs decline at a relative lower rate. This leads to a decline in agreement rate with an increase in power disparity. Further, the pattern of decline is such that it generates a threshold pattern in the assessment of power disparities from the efficiency and fairness perspectives. Low power disparities are efficient and fair whereas high power disparities are inefficient and unfair.[18]

4.2. Treatment $\mathcal{CU}$

Treatment $\mathcal{CU}$ allows the responder to choose whether or not to enter a bargain in order to explicitly account for the initial consent to initiate a relationship. Figure 9 shows the patterns of variation in mean values of agreement, disagreement, and total surplus in treatment $\mathcal{CU}$ based on the offers by all proposers and the MAOs of those responders who choose to enter. According to mean comparisons, disagreement surplus is a bad at every $x \geq 23$, and each such power disparity is subjectively inefficient.

[18] As described in Section 2.3, this result cannot be rationalized by most models of other-regarding preferences since they predict power disparities beyond a threshold will be fair and the highest power disparity will be efficient.

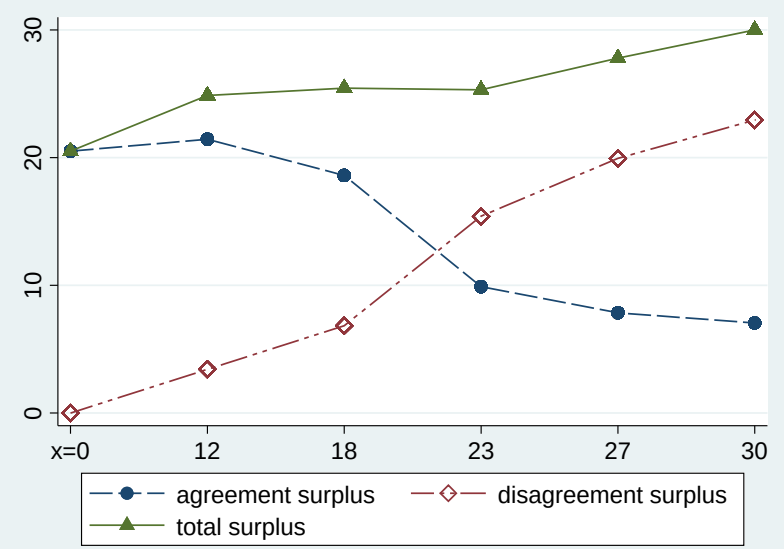

Figure 9: *Mean surplus in treatment* $\mathcal{CU}$. For each $x \in X$, these surplus values account for only those first movers who choose to enter the $\mathcal{U}(x)$ subgame. The pattern of variation in each of the three mean surplus values in treatment $\mathcal{CU}$ is largely similar to that in treatment $\mathcal{U}$.

The inferences based on comparisons of mean surplus across different values of x are generally supported by the statistical tests (see Table 2; and, Table 10 in Appendix 2 for details). However, we cannot reject the hypothesis of no difference in total surplus at any pair of power disparities in $\{12, 18, 23\}$. Hence, $x = 18$ and $x = 23$ may be objectively inefficient relative to $x = 12$ which has relatively greater agreement surplus and relatively lower disagreement surplus. Again, given that we treat a power disparity as efficient unless it can be proven inefficient, we have the following result.

Result [3] *In treatment* $\mathcal{CU}$, $x \in \{0, 12, 18\}$ *is efficient and* $x \in \{23, 27, 30\}$ *is inefficient.*

Remark 2. $x = 18$ *may possibly be objectively inefficient relative to* $x = 12$ *in both treatments because (i) agreement surplus is statistically higher and disagreement surplus is statistically lower at* $x = 12$ *than at* $x = 18$*, and (ii) total surplus values at these two power disparities are not statistically different.*

Turning to whether any efficient power disparity is unfair, we first note that no $x \geq 23$ is dispute-proof or selfish-proof according to mean comparisons of the relevant variables (Figures 10 and 11). As reported in Table 2, statistical tests confirm that every $x \geq 23$ is neither dispute-proof nor selfish-proof (see Tables 11 and 12 in Appendix 2 for details). Thus, as in treatment $\mathcal{U}$, we have the following result.

Result [4] *No efficient power disparity in* X *is unfair in treatment* $\mathcal{CU}$.

Table 2: Tests of difference for treatment $\mathcal{CU}$

Null hypothesis	**Power disparity** x					
	0	12	18	23	27	30
NBAD: S_d is not a 'bad' at x						
Signed-Rank test	–	0.01	0.01	**0.03**	**0.01**	**0.01**
DP: x is dispute proof						
Regression test	0.15	0.03	0.23	**0.01**	**0.01**	**0.01**
Fligner-Policello test	0.07	0.04	0.29	**0.01**	**0.01**	**0.01**
Kolmogorov-Smirnov test	0.11	0.01	0.48	**0.01**	**0.01**	**0.01**
SP: x is selfish proof						
Regression test	0.89	0.14	0.67	**0.01**	**0.01**	–
Fligner-Policello test	0.80	0.13	0.44	**0.01**	**0.01**	–
Kolmogorov-Smirnov test	0.99	0.39	0.83	**0.01**	**0.01**	–

Notes. The table reports *two-tailed* p-values for the null hypothesis of equality between the two relevant variables for each test. NBAD is based on the distributions of agreement and disagreement surplus across proposers (n = 55). DP is based on the distributions of offers (55 proposers) and MAOs of responders who choose IN ($\{39, 39, 37, 21, 13, 15\}$ responders at $x \in \{0, 12, 18, 23, 27, 30\}$, respectively). SP is based on the distributions of offers by the 21 other-regarding and the 34 self-regarding proposers. A proposer with a strictly positive (zero) offer in $\mathcal{U}(30)$ is classified other-regarding (self-regarding). Entries are indicated in bold when the difference is significant at the 10% level *and* the sign of the difference (test statistic) is in the direction required by the corresponding definition. p-values strictly less than 0.01 are rounded off to 0.01.

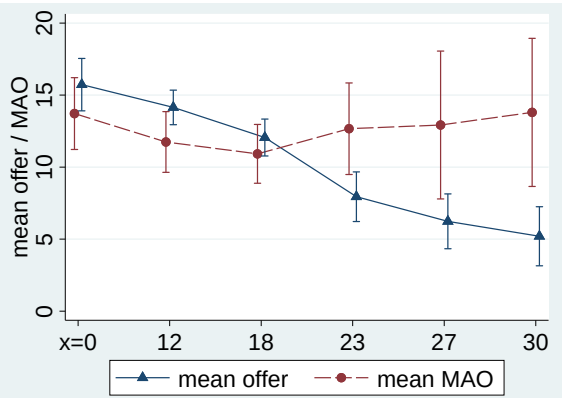

Figure 10: *Dispute proofness in treatment $\mathcal{CU}$.* As power disparity increases, mean proposer offer declines while mean MAO of responders who enter exhibits a non-monotonic pattern.

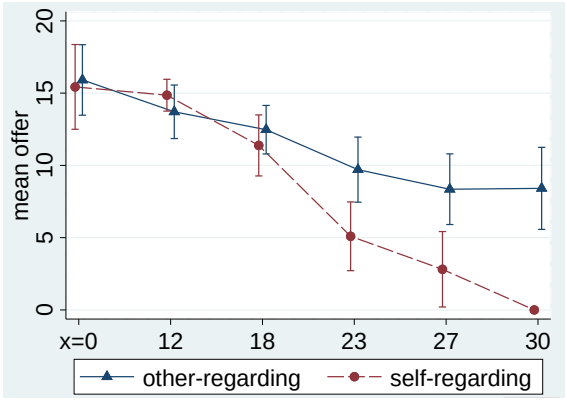

Figure 11: *Selfish proofness in $\mathcal{CU}$.* As power disparity increases, mean offers by both types of proposers decline but the rate of decline is relatively lower for other-regarding proposers.

Overall, the results do not support the presumption that efficiency is more tolerant towards power disparities than fairness. Values of $x \leq 18$ seem to be efficient as per $\mathbf{D_E}$, and we find no evidence to suggest any such power disparity is unfair.

4.3. Robustness

Given the nature of our inquiry, it is worth distinguishing between robustness checks with respect to the *primitives* versus the *operationalizations*. Our primitives are the normative commitments embodied in the two definitions: the distinction between agreement and disagreement surplus in case of efficiency, and the emphasis on accounting for perspectives of both weaker and stronger parties in case of fairness. As these form the basis of our work, we do not consider robustness checks that involve alternative normative commitments.

Our operational definition of efficiency is already quite conservative and labels some power disparities efficient that may plausibly be inefficient (e.g., $x = 18$, see Remark 2). Given our goal is to test whether any efficient power disparity is unfair, it would be uninformative to consider alternative definitions of efficiency that are *less* tolerant of power disparities than $\mathbf{D_E}$. In contrast, the operational definition of fairness is not intended to be conservative and may unduly label some fair power disparities unfair (Section 2.2.2). It relies upon the standard of reasonable expectations of weaker parties and the standard of reasonable conduct by stronger parties.

Who is reasonable enough to serve as a standard-setter is debatable. In conceptualizing reasonableness, legal scholarship typically covers a whole spectrum ranging from tolerant standards based on common practice to demanding standards that seek to capture some notion of common morality (Miller and Perry, 2012 and 2013). The basic idea being that the standard should neither be vacuous because it is unachievable, nor hostage to the percentage of people who care solely about their own material gains and losses. The following subsection considers some alternative standards that are likely to make it easier to label power disparities unfair, and thereby help assess the robustness of the results reported above. All the alternative standards seek to *exclude* self-regarding agents from serving as standard-setters. They differ in terms of the identification strategy used to classify agents as either self-regarding or other-regarding.

4.3.1. Alternative ways to assess unfair power disparities

A power disparity is unfair if it is neither dispute-proof nor selfish-proof. Our baseline test of dispute-proofness, DP, treats *all* the responders as standard-setters. We consider an alternative test, denoted DP_1, that excludes self-regarding responders from serving as standard-setters.[19] The-

[19]It is unclear whether a responder who is willing to accept anything should be considered a standard-setter while constructing the standard of reasonable expectations of the weaker parties. The case for excluding self-regarding responders seems to have been made in *Earl of Chesterfield v Janssen* (28 Eng Rep 82, 100, 1751) which has shaped the discourse on the modern doctrine of unconscionability in contract law. The court noted that an unconscionable bargain is one "such as no man in his senses and not under a delusion would make on the one hand, and as no honest

oretically, the MAO of a self-regarding responder who cares solely about her own material payoffs cannot exceed *one* at any value of x. We utilize MAOs at $x = 0$ and classify the responders with $MAO \leq 1$ as self-regarding. The MAOs of the remaining responders are used to construct the standard of reasonable expectations under DP_1.

Our baseline test of selfish-proofness, SP, treats other-regarding proposers as the standard-setters and relies on a specific identification strategy: proposers who offer any strictly positive amount at $x = 30$ are classified other-regarding. We consider two alternative ways to classify proposers. Under the first alternative, a proposer must offer at least *two* at $x = 30$ to be classified other-regarding. The resulting alternative test of selfish-proofness, denoted SP_1, may be viewed purely as a stress test or as an attempt to account for the possibility that some proposers may be particularly averse to giving nothing. The second alternative, denoted SP_2, utilizes the dictator game in treatment $\mathcal{U}$ and the trust game in treatment $\mathcal{CU}$. Proposers who offer a strictly positive amount as a dictator (trustee) in the dictator game (trust game) are labeled other-regarding.[20]

The identification strategy determines where the tests are meaningful (Table 3). For either treatment, the two tests of dispute-proofness and the three tests of selfish-proofness can be combined in a total of six ways to test whether a power disparity is unfair. The (DP, SP) combination corresponds to the baseline test and the remaining five combinations provide alternative tests. Recall that our central finding was that every $x \leq 18$ is efficient and fair in both treatments, where fairness was assessed according to the baseline (DP, SP) combination. Robustness of this result hinges on whether or not any efficient $x \leq 18$ turns out to be unfair under the five alternative tests.

Result 5. *The alternative tests contradict the fairness classification provided by the baseline test in treatment $\mathcal{U}$ in two cases: $x = 18$ is fair under the baseline test, but unfair under two out of the five alternative tests. However, these two contradictions arise under the Fligner-Policello test but not under the Kolmogorov-Smirnov test. No contradiction arises in treatment $\mathcal{CU}$.*

The results for $x = 18$ are reported in Table 4 and the rest are provided in Table 13 in Appendix 2. There exists some evidence that $x = 18$ is unfair. However, since $x = 18$ may potentially be objectively inefficient as well (Remark 2 in Section 4.2), it may be reasonable to interpret Results 1 to 5 as providing no *robust* evidence to conclude that any power disparity in $X = \{0, 12, 18, 23, 27, 30\}$ which is efficient as per $\mathbf{D_E}$ is unfair as per $\mathbf{D_F}$. Of course, given the discreteness of X, we cannot rule out that some power disparities outside X might be efficient but unfair. At the minimum, we are led to question the obviousness and the generality of the presumption that at least some efficient power disparities are unfair.

---

man would accept on the other."

[20]The amount given by proposers when acting as dictators in the dictator game in treatment $\mathcal{U}$ is statistically greater than their offers when acting as proposers offers in $\mathcal{U}(30)$ (Wilcoxon Signed-Rank $p < 0.01$). Similarly, the amount returned by proposers when acting as trustees in the trust game in treatment $\mathcal{CU}$ is statistically greater than their offers when acting as proposers in $\mathcal{CU}(30)$ (Wilcoxon Signed-Rank $p < 0.01$).

Table 3: Summary of tests for (un)fair power disparities

	What is judged?	**Standard of judgement?**
Dispute-proofness	*Offers at x by*	*MAOs at x by*
DP at each x	All proposers	All responders
DP_1 at each x	All proposers	Responders with $MAO \geq 2$ at $x = 0$
Selfish-proofness	*Offers at x by*	*Offers at x by*
SP at $x < 30$	Proposers who offer 0 at $x = 30$	Proposers who offer ≥ 1 at $x = 30$
SP_1 at $x < 30$	Proposers who offer ≤ 1 at $x = 30$	Proposers who offer ≥ 2 at $x = 30$
SP_2 at each x	Proposers who offer 0 in $\mathcal{D}/\mathcal{T}$ game	Proposers who offer ≥ 1 in $\mathcal{D}/\mathcal{T}$ game

Notes. For either treatment, the two tests of dispute-proofness – DP and DP_1 – and the three tests of selfish-proofness – SP, SP_1, and SP_2 – can be combined in a total of six ways to test whether a power disparity is unfair. The (DP, SP) combination is the baseline test used in the main analysis in Sections 4.1 and 4.2.

Table 4: Robustness check: Is $x = 18$ fair or unfair?

Combination	**Statistical test**	**Treatment $\mathcal{U}$**		**Treatment $\mathcal{CU}$**	
		p-values	Classification	p-values	Classification
(DP, SP)	F-P	(0.68, 0.05)	**Fair**	(0.29, 0.44)	**Fair**
	K-S	(0.93, 0.35)	**Fair**	(0.48, 0.83)	**Fair**
(DP, SP_1)	F-P	(0.68, 0.14)	Fair	(0.29, 0.02)	Fair
	K-S	(0.93, 0.11)	Fair	(0.48, 0.08)	Fair
(DP, SP_2)	F-P	(0.68, 0.01)	Fair	(0.29, 0.87)	Fair
	K-S	(0.93, 0.05)	Fair	(0.48, 0.99)	Fair
(DP_1, SP)	F-P	(0.09, 0.05)	**Unfair**	(0.59, 0.44)	Fair
	K-S	(0.60, 0.35)	Fair	(0.94, 0.83)	Fair
(DP_1, SP_1)	F-P	(0.09, 0.14)	Fair	(0.59, 0.02)	Fair
	K-S	(0.60, 0.11)	Fair	(0.94, 0.08)	Fair
(DP_1, SP_2)	F-P	(0.09, 0.01)	**Unfair**	(0.59, 0.87)	Fair
	K-S	(0.60, 0.05)	Fair	(0.94, 0.99)	Fair

Notes. Entries under the p-values columns are for *two-tailed* tests of dispute-proofness and selfish-proofness under the combination in the corresponding row. $x = 18$ is classified unfair when *both* dispute-proofness and selfish-proofness can be rejected at the 10% significance level. The two statistical tests are Fligner-Policello (F-P) and Kolmogorov-Smirnov (K-S). The definitions of the baseline and the five alternative combinations are provided in Table 3.

4.3.2. Internal consistency checks

We conclude the analysis by investigating two further issues which relate to the interpretation of our findings. We have used observed behavior to infer the preference type of subjects in two instances. First, we categorize proposers as either self-regarding or other-regarding based on their offers at $x = 30$ so that we can test for selfish-proofness of power disparities. Second, in rationalizing the observed patterns in responder MAOs we posited that the majority of responders have expressive preferences. A common concern in identifying preference types via differences in observed actions is that such differences may instead be driven by systematic differences in beliefs. Differences in beliefs may arise due to a host of factors including a misunderstanding of the experimental instructions and the strategic environment (Cason and Plott, 2014).

In treatment $\mathcal{U}$ we elicit at each $x \in X$ the belief of a proposer about the likelihood that his offer would be accepted by the responders (see Section 3). Figure 8 and associated results from Table 1 show that other-regarding proposers offer a significantly higher amount than self-regarding proposers at every $x \geq 23$. If both other-regarding and self-regarding proposers report, on average, similar beliefs about the likelihoods of acceptance of their offers, then we would have strong evidence that our identification has failed to capture differences in preferences and is instead capturing systematic differences in beliefs across the two groups. If the relatively higher offers of the other-regarding proposers are instead driven by differences in preferences, then we would expect to observe that they belief in a relatively higher likelihood of acceptance at $x \geq 23$. We find strong support for this: when offers by proposers that we classify other-regarding are higher than proposers that we classify self-regarding, then their beliefs about the likelihood that their offers will be accepted are also relatively higher (see Figure 12 and Table 5). This evidence is consistent with differences in behavior of self-regarding and other-regarding proposers being driven by differences in their preferences. It reassures us that the standard of reasonable conduct utilized in testing selfish-proofness of power disparities is capturing what we intend it to capture.

In treatment $\mathcal{U}$ we elicit at each $x \in X$ the belief of a responder about the likelihood that her MAO would be satisfied by the proposers. Expressive responders report a higher MAO than non-expressive responders at all values of x.[21] If both expressive and non-expressive responders report similar likelihoods of satisfaction of their MAOs, then we would have strong evidence that our identification fails to capture differences in preferences. If the higher MAOs of expressive responders are indeed driven by differences in preferences, then we would expect to observe that they attach a relatively lower likelihood to the satisfaction of their MAOs. We find expressive responders have systematically lower beliefs than non-expressive responders about proposers offers satisfying their MAOs at every power disparity (see Figure 13 and Table 5). This is consistent with the differences in MAOs of expressive and non-expressive responders being driven by differences in their preferences, and not by any misunderstanding on part of some subjects.

[21] The p-values for Fligner-Policello tests are no more than 0.05 at any $x < 30$; $x = 30$ is different by construction.

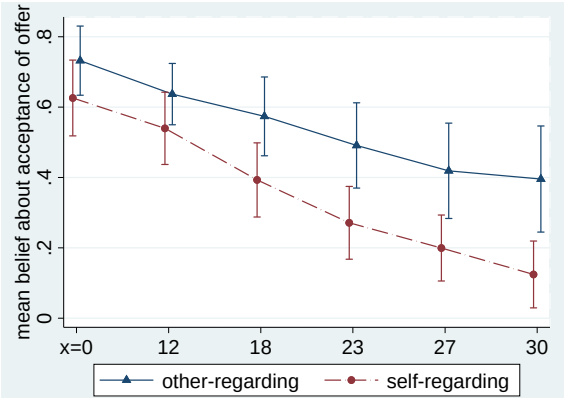

Figure 12: *Beliefs of proposers in treatment* $\mathcal{U}$. Mean belief of other-regarding proposers that their offers will be accepted by the responders is relatively greater than the corresponding mean belief of self-regarding proposers.

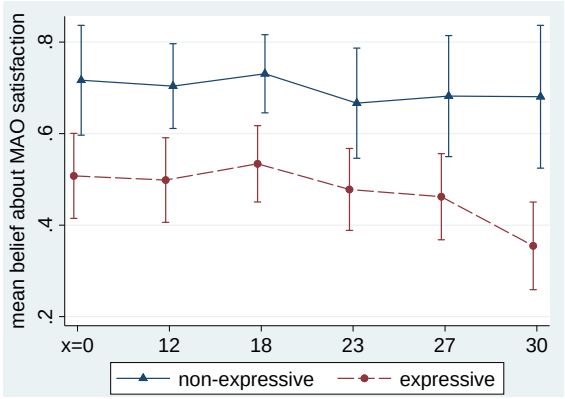

Figure 13: *Beliefs of responders in treatment* $\mathcal{U}$. Mean belief of expressive responders that their MAOs will be satisfied by proposers is relative lower than the corresponding mean belief of non-expressive responders at every power disparity.

Table 5: Tests of differences in beliefs in treatment $\mathcal{U}$

Belief comparison	**Power disparity** x					
	0	12	18	23	27	30
Other vs Self regarding proposers						
Fligner-Policello test	0.18	0.27	**0.03**	**0.01**	**0.01**	**0.01**
Kolmogorov-Smirnov test	0.77	0.33	0.23	**0.05**	**0.02**	**0.01**
Expressive vs non-expressive responders						
Fligner-Policello test	**0.01**	**0.01**	**0.01**	**0.01**	**0.01**	**0.01**
Kolmogorov-Smirnov test	**0.01**	**0.01**	**0.01**	**0.02**	**0.01**	**0.01**

Notes. Each proposer reports his belief at each x about the number of responders in the session who would accept his offer at x. A proposer with a strictly positive (zero) offer at $x = 30$ is categorized other-regarding (self-regarding). The test for proposers compares the beliefs reported by the 27 other-regarding and the 33 self-regarding proposers thus identified. Each responder reports his belief at each x about the number of proposers in the session whose offers at x satisfy her MAO at x. A responder with MAO ≥ 2 (MAO < 2) at $x = 30$ is categorized expressive (non-expressive). The test for responders compares the beliefs reported by the 35 expressive and 25 non-expressive responders thus identified. The reported two-tailed p-values correspond to a test of the null hypothesis of no difference in beliefs across the two types. Entries appear in bold if the difference is significant at the 10% level. p-values strictly less than 0.01 are rounded-off to 0.01.

5. Discussion

We find no robust evidence to support the presumption that consent-based economic efficiency is more tolerant towards power disparities than two-sided fairness. In particular, sufficiently high levels of bargaining power disparity are not only unfair but also inefficient. In the following we first draw upon some existing findings from bargaining and contracting games to conjecture that two aspects related to the notion of power disparity – its *source* and the *stage* of relationship at which it is exercised – may be crucial in evaluating power disparities. Subsequently, we relate our work to the broader debate about fairness versus efficiency in legal scholarship.

5.1. Source and stage

Our finding that sufficiently high power disparities are both inefficient and unfair should not be interpreted as providing a general and acontextual justification for intervention in environments characterized by large power disparities. Consider Fischbacher et al. (2009) and Guth et al. (1997) that study the standard ultimatum game, $\mathcal{U}(0)$, with competition among responders. They vary the *short-side* power of proposers across treatments by matching each proposer with $n \geq 1$ responders who compete to transact with the single proposer. A proposer transacts with the responder who demands the lowest share of the surplus. Since surplus cannot be generated without agreement in this setting, our definition of efficiency has no real bite and one may use total surplus (or, equivalently, agreement rate) as an operational measure of efficiency.

As in our study, average offer by proposers and the average MAO of responders decline with an increase in bargaining power disparity, as captured via n. However, unlike our study, the average MAO of responders remains below the average offer by the proposers, and both converge towards predictions of outcome-based models of other-regarding preferences as n increases. The agreement rate increases with an increase in n. These patterns suggest that if total surplus is used as the measure of efficiency, then efficiency increases with an increase in n. In addition, two-sided fairness would not consider even high values of n unfair because they are dispute-proof.[22] These contrasts with our findings suggest *as if* weaker parties perceive the ex-ante exercise of short-side power by the stronger party to be more legitimate than the power to extract profit once they are locked in a bilateral relationship.

A subtle demonstration of the perceived legitimacy of ex-ante exercise of short-side power is provided by Fehr et al. (2011). Their experiment tests whether contracts formed in a competitive setting can serve as reference points that agents perceive to be legitimate and fair, which is a central assumption in Hart and Moore's (2008) theory of "contracts as reference points." In their 'competition' treatment, two sellers participate in an English auction to earn the right to transact with a buyer. Subsequently, the matched seller chooses whether to provide 'normal' service to

[22]Given that the proposer obtains an increasing fraction of the pie with an increase in n, sufficiently high values of n may however be unfair on grounds of distributive fairness

the buyer at a low material cost to herself or 'low' service at a high material cost. In their 'random' treatment, an exogenous random device draws the auction outcome for the sellers from the distribution of auction outcomes observed in the competition treatment. The seller with the lower price transacts with the buyer in both treatments.[23]

Matched sellers rarely provide low service in the competition treatment and are significantly more likely to provide low service in the random treatment. Crucially, the treatment difference in provision of low service holds even after controlling for the transaction price. Fehr et al (2011) interpret this treatment difference as revealing the legitimizing force of ex-ante competition.

A natural question is why ex-ante competition has this legitimizing force. Perhaps the fact that sellers can ex-ante compete ensures they do not feel marginalized from the process that determines the terms of the contract even though the competition between them puts them at a power disadvantage vis-a-vis the buyer. This also helps rationalize the contrast between our findings and those of Fischbacher et al (2009) and Guth et al (1997). Power disparity that can be exploited to extract profits without genuine consent from a locked-in party thus seems to generate a substantially different behavioral response compared to the exploitation of short-side power. Perceptions of the legitimacy of power disparity thus seem sensitive to the source of power disparity (here, x vs n) and the *stage* of the relationship at which it can be exercised (pre vs post lock-in).

Attending to the source and stage of power disparity may allow us to better answer *when* does power disparity warrant intervention and on *what* grounds. Section 208(d) of the Restatement (Second) of Contracts (1981) states that "a bargain is not unconscionable merely because the parties to it are unequal in bargaining position". The default position of no intervention despite bargaining power disparity can be viewed as the efficiency guided view that voluntary transactions make all involved parties better-off. From this default position, several exceptions arise. Consistent with this efficiency-guided view, regulatory bodies may use evidence of market failure or anti-competitive conduct to justify intervention. In contrast, present day contract law contains many "broad standards based on reliance, reasonableness, and good faith ... that can be understood in fairness terms" to deal with issues arising in ongoing contractual relationships (Scott, 2004). Our results provide some empirical support to justify these standards on grounds of two-sided fairness. Perhaps more importantly, they suggest that these standards may also be justified on grounds of consent-based efficiency even in the absence of market failures. The next subsection discusses how our results sit within the broader debate between efficiency and fairness in the law and economics literature.

5.2. Efficiency, fairness, and law

The rise of the law and economics movement following the seminal contributions of Coase (1960), Calabresi (1961), and Posner (1973) suggested that the concept of economic efficiency may provide

[23]The actual design of Fehr et al (2011) is more elaborate but this caricature suffices to convey our point.

a positive explanation for the evolution of (common) law and serve as a useful normative guide in shaping the law. Posner (1973, 1979, 1980) proposed 'social wealth maximization' as an operational definition of economic efficiency, which is achieved when resources lie in hands of those who value them most.[24] A common critique of this operational definition was that it equates 'value' with the willingness *and* the ability to pay. Dworkin (1980) criticized it for recommending a planner to avoid transaction costs and increase a society's wealth by *forcibly* allocating resources to users who value them most. Dworkin's critique may be viewed as highlighting that social wealth maximization is inherently contrary to the normative idea of consent that underpins any sensible notion of economic efficiency. Accounting for these critiques some scholars summarized that "[T]he normative theory of efficiency is relatively uncontroversial (Who favors wasting money?) as a broad guide to policy. But, controversy is abundant when efficiency is seen as dominating other norms of fairness and justice" (Cooter and Rubinfeld, 1989, pp. 1068).

The next and as yet the most comprehensive attempt at formalizing the tension between economic efficiency and fairness has been undertaken by Kaplow and Shavell (2001a, 2001b). They define an efficient choice as one that maximizes some social welfare function that depends *solely* on individual's well-being, where an individual's well-being depends on everything that the individual herself deems valuable (and may thus include her personal taste for fairness). Any function that departs from the abovementioned class is labeled a fairness-based welfare function. Under these definitions it can be proven that a policy choice dictated by notions of fairness can be Pareto dominated. Kaplow and Shavell therefore recommend that policy choices should not be driven by fairness considerations. They emphasize that while this conclusion follows almost tautologically from their definitions it is nonetheless useful because "the depth of the tension between fairness and welfare is not widely appreciated". Even this work has at best received "mixed-success in persuading non-economically-oriented legal scholars" that welfare maximization ought to be the singular goal of legal policymaking (Hermalin et al., 2006). In addition to the practical concerns – unobservability of utility functions, the need to conduct interpersonal comparisons of utility, and the lack of a compelling justification for choosing a particular welfare function – it is driven by the longstanding divide between consequentialist and deontological ethics (see Coleman, 2003).[25]

The evolution of contract law attests to the depth of the tension between efficiency and fairness (Horwitz, 1974; Edwards, 2009). In resolving contractual disputes and filling contractual gaps

[24]It must be remarked that a unique compelling operationalization of economic efficiency is challenging. Consider, for instance, the debate about whether or not common law rules evolve towards efficiency. There exists a voluminous theoretical literature on this question (see Gennaioli and Shleifer, 2007). Empirical assessments of this question, however, have to rely on such operational measures of efficiency that are at best proxies for voluntariness of the underlying interactions or the welfare of agents. For example, Niblett et al. (2010) focus on a particular legal rule and consider three operationalizations that equate efficiency with *variability* in the application of the legal rule across different jurisdictions over time in slightly different ways.

[25]This divide is more than a philosophical curiosity and impacts how courts resolve legal disputes. Mainali et al. (2018) utilize computational linguistics and machine learning to classify US Circuit Court Opinions as revealing a consequentialist or deontological bent, quantify the evolution of this bent over a century across states in USA, and find that it is strongly correlated with the law school attended by judges.

nineteenth century contract law typically utilized bright line rules and favored a textual interpretation of written contracts. The increased complexity of contractual relationships over time and the inevitable incompleteness of contracts led to a shift in how courts go about resolving contractual disputes. By mid-twentieth century, the approach based on rigid rules and textual interpretation started giving way to one based on broad standards and contextual interpretation (Kennedy, 1976; Speidel, 1982; Fried, 2015). Today, a reference to 'fairness' and related notions can be found in enough provisions of contract law, often in relation to the potential abuse of bargaining power, that legal scholars largely agree that "[T]he principle of fairness is entrenched in legal doctrine, including contract doctrine. Equitable estoppel, quantum meruit, unjust enrichment, the doctrine of avoidable consequences, unconscionability, good faith, reasonableness, and reformation are just a few of the contract doctrines that can be understood in fairness terms" (Scott, 2004 pp. 382). Taken together, they serve as background guidelines under which agents may freely contract, and help courts resolve disputes relating to procedural or substantive issues in the formation or execution of contracts.

Our definition of an economically efficient power disparity is rooted in mutual consent and does not rely on our perceptions of distributive fairness. It circumvents the concerns arising from unobservability of individual utility functions and interpersonal comparisons of utility. Our definition of an unfair power disparity operationalizes notions of reasonable expectations and reasonable conduct that motivate many of the abovementioned doctrines. Our work, of course, does not provide any guidance for how to resolve specific contractual disputes. As highlighted in Section 5.1, it nonetheless suggests that the existence of these provisions need not be grounded only in fairness and may be rationalized by consent-based efficiency as well.

6. Conclusion

The paper investigates the potential tension between efficiency and fairness in demarcating permissible and impermissible bargaining power disparities. Specifically, we test the presumption that economic efficiency is more tolerant towards power disparities than fairness. We utilize simple variants of the well researched ultimatum game that allow us to distinguish between surplus realized with or without mutual consent between bargaining parties. We propose a definition to categorize power disparities as economically efficient or inefficient by assuming mutual consent is the normative principle behind the idea of economic efficiency. The difficulty of uniquely defining an unfair power disparity needs little elaboration. Hence, we simply draw upon legal scholarship and judicial practice to propose a two-sided definition of an unfair power disparity that tries to accommodate the perspectives of both weaker and stronger bargaining parties.

Our work is rooted in the revealed preference approach and the definitions do not label any level of power disparity unfair or inefficient, *a priori*. Objective inefficiency seems normatively compelling but provides little help in actually demarcating permissible and impermissible power disparities in our data. This seems to support the view that "the acceptability of a moral principle

is inverse to its capacity to resolve an actual issue" (Posner, 1998). The efficiency classification of power disparities in our data is driven by subjective inefficiency which embodies a heightened and perhaps less orthodox concern for mutual consent in the realization of surplus. Further, the demarcation provided by subjective inefficiency is similar to that provided by two-sided fairness.

We conclude by noting that the law is routinely called upon to judge whether one party sought a profit without the consent of another despite the difficulty of unambiguously defining and ascertaining consent. However, in doing so, mutual consent is often viewed as a marker of fairness, and not necessarily efficiency. For instance, Section 208(d) of the Restatement (Second) of Contracts (1981) states that "... gross inequality of bargaining power, together with terms unreasonably favorable to the stronger party ... may show that the weaker party had no meaningful choice, no real alternative, or did not in fact assent or appear to assent to the unfair terms." To an economist, in contrast, mutual consent would likely be the prime marker of voluntariness and economic efficiency. Viewed in this light, the observed correspondence between consent-based efficiency and two-sided fairness in the evaluation of power disparities may not be surprising. What is surprising is that there now exists a voluminous theoretical and empirical literature on 'fairness' within economics but not on 'mutual consent'.

References

Acemoglu, Daron and Alexander Wolitzky. 2011. "The Economics of Labor Coercion." *Econometrica*, 79(2): 555-600.

Anbarci, Nejat and Nick Feltovich. 2013. "How Sensitive are Bargaining Outcomes to Changes in Disagreement Payoffs?" *Experimental Economics*, 16 (4): 560-596.

Basu, Kaushik. 2007. "Coercion, Contract and the Limits of the Market." *Social Choice and Welfare*, 29: 559-579.

Berg, Joyce, John Dickhaut, and Kevin McCabe. 1995. "Trust, Reciprocity, and Social History". *Games and Economic Behavior*, 10(1): 122-142.

Binmore, Ken, Peter Morgan, Avner Shaked, and John Sutton. 1991. "Do People Exploit their Bargaining Power." *Games and Economic Behavior*, 3(3): 295-322.

Bowles, Samuel and Herbert Gintis. 1992. "Power and Wealth in a Competitive Capitalist Economy". *Philosophy and Public Affairs*, 21(4): 324-353.

Calabresi, Guido. 1961. "Some Thoughts on Risk Distribution and the Law of Torts". *Yale Law Journal*, 70(4): 499-553.

Cason, Timothy N. and Charles R. Plott. 2014. "Misconceptions and Game Form Recognition: Challenges to Theories of Revealed Preference and Framing." *Journal of Political Economy*, 122(6): 1235-70.

Chemerinsky, Erwin. 2016. "The Rational Basis Test is Constitutional (and Desirable)". *Georgetown Journal of Law & Public Policy*, 14: 401-419.

Coase, Ronald H. 1960. "The Problem of Social Cost". *Journal of Law & Economics*, 3(Oct): 1-44.

Coleman, Jules L. 2003. "The Grounds of Welfare". *Yale Law Journal*, 112(6): 1511-1543.

Cooper, David and John Kagel. 2016. "Other-regarding Preferences". In *The Handbook of Experimental Economics*, Volume 2, pp. 217-289. Ed. by Kagel, John and Alvin Roth. Princetion University Press.

Cooter, Robert D. and Daniel L. Rubinfeld. 1989. "Economic Analysis of Legal Disputes and their Resolution." *Journal of Economic Literature*, 27(3): 1067-1097.

Craswell, Richard. 1993. "Property Rules and Liability Rules in Unconscionability and Related Doctrines." *University of Chicago Law Review*, 60(1): 1-65.

Davidov, Guy. 2016. *A Purposive Approach to Labour Law*. Oxford University Press.

Dufwenberg, Martin, Paul Heidhues, Georg Kirchsteiger, Frank Riedl, and Joel Sobel. 2011. "Other-regarding Preferences in General Equilibrium". *Review of Economic Studies*, 78: 613-639.

Dworkin, Ronald M. 1980. "Is Wealth a Value?" *Journal of Legal Studies*, 9(2): 191-226.

Edwards, Carolyn. 2009. "Freedom of Contract and Fundamental Fairness for Individual Parties: The Tug of War Continues". *UMKC Law Review*, 77(3): 647-696.

Eisenberg, Melvin A. 1982. "The Bargain Principle and its Limits". *Harvard Law Review*, 95(4): 741-801.

Epstein, Richard A. 1975. "Unconscionability: A Critical Reappraisal". *Journal of Law & Economics*, 18(2): 293-315.

Farnsworth, Allan E. 1982. *Contracts*. Little, Brown and Company.

Fehr, Ernst and Klaus M. Schmidt. 1999. "A Theory of Fairness, Competition, and Cooperation". *Quarterly Journal of Economics*, 114(3): 817-868.

Fehr, Ernst, Oliver Hart, and Christian Zehnder. 2011. " Contracts as Reference Points – Experimental Evidence". *American Economic Review*, 101(2): 493-525.

Fischbacher, Urs. 2007. "z-Tree: Zurich Toolbox for Ready-made Economic Experiments". *Experimental Economics*, 10(2): 171-178.

Fischbacher, Urs, Christina M. Fong, and Ernst Fehr. 2009. "Fairness, Errors, and the Power of Competition". *Journal of Economic Behavior & Organization*, 72: 527545.

Fogel, Robert W. 1989. *Without Consent or Contract: The Rise and Fall of American Slavery*. W. W. Norton & Company.

Fried, Charles. 2015. *Contract as Promise: A Theory of Contractual Obligation*. Oxford University Press; Second edition.

Gennaioli, Nicola, and Andrei Shleifer. 2007. "The Evolution of Common Law". *Journal of Political Economy*, 115 (1): 43-68.

Glaeser, Edward L., and Andrei Shleifer. 2003. "The Rise of the Regulatory State". *Journal of Economic Literature*, 41(2): 401-425.

Glaeser, Edward L., Giacomo A. M. Ponzetto, and Andrei Shleifer. "Securing Property Rights". *Working Paper*.

Greiner, Ben. 2004. "An Online Recruitment System for Economic Experiments." In *Forschung Wissenschaftliches Rechnen*. Gesellschaft fiir wissenschaftliche Datenverarbeitung Bericht, Vol. 63, ed. Kurt Kremer and Volker Macho, 79-93. Gottingen, Germany: Gesellschaft ftir wissenschaftliche Datenverarbeitung.

Guth, Werner, Nagede Marchand, and Jean-Louis Rulliere. 1997. "On the Reliability of Reciprocal Fairness: An Experimental Study". Humboldt Univeristy of Berlin. Working paper.

Handgraaf, M.J.J., Van Dijk, E., Vermunt, R.C., Wilke, H.A.M., and De Dreu, C.K.W. 2008. Less Power or Powerless? Egocentric Empathy Gaps and the Irony of Having Little versus No Power in

Social Decision Making. *Journal of Personality and Social Psychology*, 95: 1136-1149.

Hart, Oliver, and John Moore. 2008. "Contracts as Reference Points." *Quarterly Journal of Economics*, 123(1): 1-48.

Hayek, Friedrich. 1960. *The Constitution of Liberty*, Chicago: University of Chicago Press.

Herbert, Alan P. 1935. *Uncommon Law.* Meuthen.

Hennig-Schmidt, Heike, Bernd Irlenbusch, Rainer M. Rilke, and Gary Walkowitz. 2018. "Asymmetric Outside Options in Ultimatum Bargaining: A Systematic Analysis". *International Journal of Game Theory*, 47: 301-329.

Hermalin, Benjamin E., Avery W. Katz, and Richard Craswell. "Contract Law". In the *Handbook of Law and Economics, Volume 1.* Edited by A.M. Polinsky and Steven Shavell. North Holland.

Hillman, Arye L. 2010. "Expressive Behaviour in Economics and Politics". *European Journal of Political Economy*, 26(4): 403-418.

Horwitz, Morton J. 1974. "The Historical Foundations of Modern Contract Law". *Harvard Law Review*, 87(5): 917-956.

Kaplow, Louis and Steven Shavell. 2001a. "Fairness versus Welfare". *Harvard Law Review*, 114(4): 961-1388.

Kaplow, Louis and Steven Shavell. 2001b. "Any Non-welfarist Method of Policy Assessment Violates the Pareto Principle". *Journal of Political Economy*, 109(2): 281-286.

Kaplow, Louis and Steven Shavell. 2003. "Fairness versus Welfare: Notes on the Pareto Principle, Preferences, and Distributive Justice". *Journal of Legal Studies*, 32(1): 331-362.

Kennedy, Duncan. 1976. "Form and Substance in Private Law Adjudication". *Harvard Law Review*, 89(8): 1685-1778.

Kessler, Freidrich. 1943. "Contracts of Adhesion: Some Thoughts about Freedom of Contract". *Columbia Law Review*, 43: 629-642.

Landes, William M. and Richard A. Posner. 1978. "Salvors, Finders, Good Samaritans, and Other Rescuers: An Economic Study of Law and Altruism." *Journal of Legal Studies*, 7(1): 83-128.

Llewellyn, Karl N. 1960. *The Common Law Tradition: Deciding Appeals*, Boston: Little, Brown and Company.

Mainali, Nischal, Liam Meier, Elliott Ash, and Daniel L. Chen. "Automated Classification of Modes of Moral Reasoning in Judicial Decisions". *Computational Legal Studies*, Forthcoming.

Miller, Alan D. and Ronen Perry. 2013. "Good Faith Performance". *Iowa Law Review*, 98: 689-745.

Miller, Alan D. and Ronen Perry. 2012. "The Reasonable Person". *New York University Law Review*, 87(2): 323-392.

Mitchell, Catherine. 2003. "Leading a Life of its Own? The Roles of Reasonable Expectation in Contract Law". *Oxford Journal of Legal Studies*, 23(4): 639-665.

Mnookin, Robert H. and Lewis Kornhauser. 1979. "Bargaining in the Shadow of the Law: The Case of Divorce". *Yale Law Journal*, 88(5): 950-997.

Moran, Mayo P. 2003. *Rethinking the Reasonable Person: An Egalitarian Reconstruction of the Objective Standard*. Oxford University Press.

Niblett, Anthony, Richard A. Posner, and Andrei Shleifer. 2010. "The Evolution of a Legal Rule." *Journal of Legal Studies*, 39(2): 325-358.

Piccione, Michele and Ariel Rubinstein. 2007. "Equilibrium in the Jungle." *Economic Journal*, 117: 883-896.

Posner, Eric. 1995. "Contract Law in the Welfare State: A Defense of the Unconscionability Doctrine, Usury Laws, and Related Limitations on the Freedom to Contract." *Journal of Legal Studies*, 24(2): 283-319.

Posner, Richard A. 1973. *Economic Analysis of Law*. New York: Little Brown and Co.

Posner, Richard A. 1979. "Utilitarianism, Economics, and Legal Theory." *Journal of Legal Studies*, 8(1): 103-140.

Posner, Richard A. 1980. "The Ethical and Political Basis of the Efficiency Norm in Common Law Adjudication." *Hofstra law Review*, 8(3): 487-507.

Posner, Richard A. 1998a. "The Problematics of Moral and Legal Theory." *Harvard Law Review*, 111(7): 1637-1717.

Posner, Richard A. 1998b. "Reply to the Critics of "The Problematics of Moral and Legal Theory"." *Harvard Law Review*, 111(7): 1796-1823.

Scalet, Steven P. 2003. "Fitting the People they are Meant to Serve: Reasonable Persons in the American Legal System". *Law and Philosophy*, 22(1): 75-110.

Schelling, Thomas C. 1981. "Economic Reasoning and the Ethics of Policy." *Public Interest*, 63: 37-61.

Schwab, Stewart J. 2017. "Law-and-Economics Approaches to Labour and Employment Law". *International Journal of Comparative Labour Law and Industrial Relations*, 33(1): 115-144.

Scott, Robert E. 2004. "The Death of Contract Law". *University of Toronto Law Journal*, 54(4): 369-390.

Speidel, Richard E. 1982. "The New Spirit of Contract". *Journal of Law and Commerce*, 2: 193-209.

Thaler, Richard H. 2016. *Misbehaving: The Making of Behavioral Economics*. W. W. Norton & Company.

Trebilcock, M.J. 1976. "The Doctrine of Inequality of Bargaining Power: Post-Benthamite Economics in the House of Lords". *University of Toronto Law Journal*, 26(4): 359-385.

Unger, Roberto M. 1983. "The Critical Legal Studies Movement". *Harvard Law Review*, 96(3): 561-675.

van Dijk, Eric, and Riel Vermunt. 2000. "Strategy and Fairness in Social Decision Making: Sometimes it Pays to be Powerless". *Journal of Experimental Social Psychology*, 36(1): 1-25.

Williamson, Oliver E. 1985. *The Economic Institutions of Capitalism*. New York: Macmillan.

APPENDICES (NOT FOR PUBLICATION)

- Appendix 1 provides the proof for Example in Section 2.3.1.
- Appendix 2 contains the the tables with empirical results referred to in the paper.
- Appendix 3 contains the experimental instructions and screenshots.

Appendix 1

Proof of Example 1. The aggregate agreement rate at $x \in [0, \pi]$ is given by $\alpha(x) = \alpha_o - \beta\frac{x}{\pi}$, where $\beta \in [0, \alpha_o]$ for any $\alpha_o \in [0, 1]$. The proof relies on two basic observations. First, since agreement rate continuously declines with an increase in x, agreement surplus will monotonically and continuously decrease and disagreement surplus will monotonically and continuously increase with an increase in x. As $S_a(x) < S_d(x)$ is a necessary condition for x to be subjectively inefficient, if $S_a(\pi) > S_d(\pi)$ then no x can be subjectively inefficient. Second, total surplus at x is $S_t(x) = (\alpha_o - \beta\frac{x}{\pi})\pi + (1 - \alpha_o + \beta\frac{x}{\pi})x$, such that $S_t'(x) = (1 - \alpha_o - \beta) + \frac{2\beta x}{\pi}$ and $S_t''(x) = \frac{2\beta}{\pi}$. Thus, total surplus will monotonically and continuously increase in x if $S_t'(0) \geq 0$. Given that agreement surplus monotonically decreases and disagreement surplus monotonically increases, no x can be objectively inefficient if $S_t'(0) \geq 0$. These two observations help demarcate the following cases.

(I) $S_a(\pi) \geq S_d(\pi)$ and $S_t'(0) \geq 0$: No x will be inefficient since necessary conditions for an x to be subjectively or objectively inefficient do not hold.

(II) $S_a(\pi) < S_d(\pi)$ and $S_t'(0) \geq 0$: Some values of x can be subjectively inefficient but no x can be objectively inefficient as an increase in total surplus is always accompanied with a decrease in agreement surplus. Due to continuity and monotonicity, there will necessarily exist a unique $x_{ad} < \pi$ such that disagreement surplus is greater (lower) than agreement surplus at every x larger (smaller) than x_{ad}. Disagreement surplus will thus be a bad at every $x > x_{ad}$. Further, relative to any $y \in [0, x_{ad}]$, every $x \in (x_{ad}, \pi]$ will have a greater total surplus which will be driven entirely by its relatively grater disagreement surplus. Hence, every $x \in (x_{ad}, \pi]$ will be subjectively inefficient relative to any $y \in [0, x_{ad}]$.

(III) $S_a(\pi) \geq S_d(\pi)$ and $S_t'(0) < 0$: Some values of x can be objectively inefficient but no x can be subjectively inefficient as disagreement surplus in not a bad at any $x \in [0, \pi]$. Total surplus is a quadratic function of x which first decreases around $x = 0$ and then increases to π at $x = \pi$. Hence, for any $\alpha_o < 1$, there will necessarily exist a unique $x_t < \pi$ such that $S_t(0) = S_t(x_t)$. Total surplus is relatively greater (lower) at every $x \in (x_t, \pi]$ ($x \in (0, x_t)$) than the total surplus at $x = 0$. Further, relative to $x = 0$, every $x \in (0, x_t]$ will have a strictly lower agreement surplus and a strictly greater disagreement surplus. Hence, every $x \in (0, x_t]$ will be objectively inefficient relative to $x = 0$.

Finally, if $S_a(\pi) < S_d(\pi)$ and $S_t'(0) < 0$, then both objective and subjective inefficiency may come into play. However, depending upon whether x_t is lower or higher than x_{ad} we can distinguish the following two cases.

(IV) $S_a(\pi) < S_d(\pi)$, $S_t'(0) < 0$, and $x_t < x_{ad}$: Note that

- relative to $x = 0$, every $x \in (0, x_t]$ will have no more total surplus, strictly lower agreement surplus, and strictly higher disagreement surplus. Hence, every $x \in (0, x_t]$ will be objectively inefficient relative to $x = 0$.
- relative to $x = 0$, every $x \in (x_{ad}, \pi]$ will have a strictly greater total surplus but this will be driven entirely by its greater disagreement surplus. Since disagreement surplus is a bad at every $x \in (x_{ad}, \pi]$, each such x will be subjectively inefficient relative to $x = 0$.

(V) $S_a(\pi) < S_d(\pi)$, $S_t'(0) < 0$, and $x_t \geq x_{ad}$: In light of case IV, every $x \in (0, x_t]$ is objectively inefficient relative to $x = 0$ and every $x \in (x_{ad}, \pi]$ is subjectively inefficient relative to $x = 0$. When $x_t \geq x_{ad}$, then every $x \in (0, \pi]$ will be objectively or subjectively inefficient relative to $x = 0$.

Appendix 2

Table 6: Tests of order effects in proposer offers and responder MAOs

	Power disparity x						**Pooled over x**	**$\mathcal{D}/\mathcal{T}$ game**
	0	12	18	23	27	30		
Proposer offers								
Treatment $\mathcal{U}$	0.97	0.11	0.31	0.72	0.92	0.95	0.66	0.63
Treatment $\mathcal{CU}$	0.28	0.16	0.48	0.21	0.56	0.46	0.90	0.65
Responder MAOs								
Treatment $\mathcal{U}$	**0.01**	0.25	0.51	0.54	0.91	0.98	0.37	

Notes. The table reports two-tailed p-values for the null hypothesis of equality in offers/MAOs when the ultimatum block is implemented before versus after the dictator/trust block under the Fligner-Policello test. Entries appear in bold if the difference is significant at the 10% significance level. We do not test for order effects in responder MAOs in treatment $\mathcal{CU}$ because the sample sizes for each order are small due to the endogenous choice of entry.

Table 7: Tests of differences in surplus at pairs of power disparities in treatment $\mathcal{U}$

	Power disparity x					
	0	12	18	23	27	30
Total Surplus						
0						
12	0.01^{+}					
18	0.01^{+}	0.17^{+}				
23	0.01^{+}	0.04^{+}	0.01^{+}			
27	0.01^{+}	0.01^{+}	0.01^{+}	0.01^{+}		
30	0.01^{+}	0.01^{+}	0.01^{+}	0.01^{+}	0.01^{+}	
Agreement Surplus						
0						
12	0.78^{-}					
18	0.01^{-}	0.01^{-}				
23	0.01^{-}	0.01^{-}	0.01^{-}			
27	0.01^{-}	0.01^{-}	0.01^{-}	0.21^{-}		
30	0.01^{-}	0.01^{-}	0.01^{-}	0.23^{-}	0.26^{-}	
Disagreement Surplus						
0						
12	0.01^{+}					
18	0.01^{+}	0.01^{+}				
23	0.01^{+}	0.01^{+}	0.01^{+}			
27	0.01^{+}	0.01^{+}	0.01^{+}	0.01^{+}		
30	0.01^{+}	0.01^{+}	0.01^{+}	0.01^{+}	0.01^{+}	

Notes. The table reports *two-tailed* p-values for the null hypothesis of equality between the two relevant variables under the Signed-Rank test. We test for differences in the distribution of surplus across a pair of power disparities. Let $S_j^p(x') - S_j^p(x)$ be the paired difference in surplus of type $j \in \{a, d, t\}$ for proposer p at power disparities x‘ and x, with $x' > x$. Rank the absolute values of the paired differences from the lowest to the highest. The superscript “+” (“-”) indicates that the sum of the ranks of the positive-valued paired differences is greater (smaller) than the corresponding sum for the negative-valued paired differences. 60 proposers. p-values strictly less than 0.01 are rounded off to 0.01.

Table 8: Tests of dispute-proofness of power disparities in treatment $\mathcal{U}$

Y = offer/MAO	**Regression test**		**F-P test**	**K-S test**
	Model 1	Model 2		
$g(0) - MAO(0)$	2.84^{*}		2.87^{**}	2.87^{*}
	(0.06)		(0.05)	(0.08)
$g(12) - MAO(12)$	1.17		1.20	1.20
	(0.28)		(0.25)	(0.66)
$g(18) - MAO(18)$	-0.21		-0.18	-0.18
	(0.83)		(0.68)	(0.92)
$g(23) - MAO(23)$	-2.93^{***}		-2.90^{***}	-2.90^{**}
	(0.01)		(0.01)	(0.03)
$g(27) - MAO(27)$	-3.48^{***}		-3.45^{***}	-3.45^{**}
	(0.01)		(0.01)	(0.02)
$g(30) - MAO(30)$	-4.90^{***}		-4.87^{***}	-4.87^{***}
	(0.01)		(0.01)	(0.01)
x		-0.21^{***}		
		(0.01)		
$\mathbb{I}_p$		3.57^{**}		
		(0.02)		
$x \cdot \mathbb{I}_p$		-0.26^{***}		
		(0.01)		
x dummies	Yes			

Notes. p-values in parentheses. *: p-value ≤ 0.1 **: p-value ≤ 0.05 ***: p-value ≤ 0.01; p-values strictly less than 0.01 are rounded off to 0.01. Columns 1 and 2 report the results of random effects regressions where the dependent variable is the offer (MAO) if the observation corresponds to a proposer (responder). Column 1 provides estimates of the coefficients of $\mathbb{I}_x \cdot \mathbb{I}_p$ at each $x \in \{0, 12, 18, 23, 27, 30\}$, which we denote $g(x) - MAO(x)$ since they can be interpreted as the difference between the average proposer offer and the average responder MAO at x. See Section 3.3 for the full specification. Standard errors in Columns 1 and 2 are clustered around individuals, with individual characteristics (gender, undergrad/postgrad, faculty, previous experience with experiments, Australia born) included as regressors. Column 3 (4) reports the results of the non-parametric Fligner-Policello (Kolmogorov-Smirnov) test of differences in proposer offers and responder MAOs at each value of x. The coefficients in Columns 3 and 4 are the raw differences in means. 60 proposers and 60 responders used for all estimates.

Table 9: Tests of selfish-proofness of power disparities in treatment $\mathcal{U}$

Y = offer/MAO	**Regression**		**F-P test**	**K-S test**
	Model 1	Model 2		
$g_s(0) - g_o(0)$	-1.75		-1.16	-1.16
	(0.37)		(0.41)	(0.58)
$g_s(12) - g_o(12)$	-1.38		-0.79	-0.79
	(0.36)		(0.73)	(0.98)
$g_s(18) - g_o(18)$	-4.17^{***}		-3.58^{**}	-3.58
	(0.01)		(0.05)	(0.35)
$g_s(23) - g_o(23)$	-5.68^{***}		-5.09^{***}	-5.09^{**}
	(0.01)		(0.01)	(0.03)
$g_s(27) - g_o(27)$	-6.93^{***}		-6.34^{***}	-6.34^{***}
	(0.01)		(0.01)	(0.01)
$g_s(30) - g_o(30)$			-7.00	-7.00
			(†)	(†)
x		-0.36^{***}		
		(0.01)		
$\mathbb{I}_p$		-0.72		
		(0.72)		
$x \cdot \mathbb{I}_p$		-0.20^{**}		
		(0.05)		
x dummies	Yes			

Notes. p-values in parentheses. *: p-value ≤ 0.1 **: p-value ≤ 0.05 ***: p-value ≤ 0.01; p-values strictly less than 0.01 are rounded off to 0.01. †: different by construction. Columns 1 and 2 report results of random effects regressions where the dependent variable is the proposer's offer. Column 1 provides estimates of the coefficients of $\mathbb{I}_x \cdot \mathbb{I}_{SRP}$ at each $x \in \{0, 12, 18, 23, 27, 30\}$, which we denote $g_s(x) - g_o(x)$ since they can be interpreted as the difference in the average offer between self-regarding and other-regarding proposers at x. See Section 3.3 for the full specification. Standard errors in Columns 1 and 2 are clustered around individuals, with individual characteristics (gender, undergrad/postgrad, faculty, previous experience with experiments, Australia born) included as regressors. Column 3 (4) reports the results of the non-parametric Fligner-Policello (Kolmogorov-Smirnov) test of differences in offers between self-regarding and other-regarding proposers at each x. The coefficients in Columns 3 and 4 are the raw differences in means. 27 other-regarding and 33 self-regarding proposers used for all estimates.

Table 10: Tests of differences in surplus at pairs of power disparities in treatment $\mathcal{CU}$

	Power disparity x					
	0	12	18	23	27	30
Total Surplus						
0						
12	0.01^{+}					
18	0.01^{+}	0.51^{+}				
23	0.01^{+}	0.93^{+}	0.68^{-}			
27	0.01^{+}	0.01^{+}	0.01^{+}	0.01^{+}		
30	0.01^{+}	0.01^{+}	0.01^{+}	0.01^{+}	0.01^{+}	
Agreement Surplus						
0						
12	0.28^{+}					
18	0.06^{-}	0.01^{-}				
23	0.01^{-}	0.01^{-}	0.01^{-}			
27	0.01^{-}	0.01^{-}	0.01^{-}	0.01^{-}		
30	0.01^{-}	0.01^{-}	0.01^{-}	0.01^{-}	0.03^{-}	
Disagreement Surplus						
0						
12	0.01^{+}					
18	0.01^{+}	0.01^{+}				
23	0.01^{+}	0.01^{+}	0.01^{+}			
27	0.01^{+}	0.01^{+}	0.01^{+}	0.01^{+}		
30	0.01^{+}	0.01^{+}	0.01^{+}	0.01^{+}	0.01^{+}	

Notes. The table reports *two-tailed* p-values for the null hypothesis of equality between the two relevant variables under the Signed-Rank test. We test for differences in the distribution of surplus across a pair of power disparities. Let $S_j^p(x') - S_j^p(x)$ be the paired difference in surplus of type $j \in \{a, d, t\}$ for proposer p at power disparities x‘ and x, with $x' > x$. Rank the absolute values of the paired differences from the lowest to the highest. The superscript “+” (“-”) indicates that the sum of the ranks of the positive-valued paired differences is greater (smaller) than the corresponding sum for the negative-valued paired differences. 55 proposers. p-values strictly less than 0.01 are rounded off to 0.01.

Table 11: Tests of dispute-proofness of power disparities in treatment $\mathcal{CU}$

Y = offer/MAO	**Regression**		**FP**	**KS**
	Model 1	Model 2		
$g(0) - MAO(0)$	2.14		2.01*	2.01
	(0.15)		(0.07)	(0.11)
$g(12) - MAO(12)$	2.44**		2.40**	2.40**
	(0.03)		(0.04)	(0.01)
$g(18) - MAO(18)$	1.29		1.14	1.14
	(0.23)		(0.29)	(0.48)
$g(23) - MAO(23)$	−3.78***		−4.72***	−4.72***
	(0.01)		(0.01)	(0.01)
$g(27) - MAO(27)$	−5.95***		−6.69***	−6.69***
	(0.01)		(0.01)	(0.01)
$g(30) - MAO(30)$	−7.58***		−8.60***	−8.60***
	(0.01)		(0.01)	(0.01)
x		−0.05		
		(0.38)		
$\mathbb{I}_p$		4.17***		
		(0.01)		
$x \cdot \mathbb{I}_p$		−0.31***		
		(0.01)		
x dummies	Yes			

Notes. p-values in parentheses. *: p-value ≤ 0.1 **: p-value ≤ 0.05 ***: p-value ≤ 0.01; p-values strictly less than 0.01 are rounded off to 0.01. Columns 1 and 2 report the results of random effects regressions where the dependent variable is the offer (MAO) if the observation corresponds to a proposer (responder). Column 1 provides estimates of the coefficients of $\mathbb{I}_x \cdot \mathbb{I}_p$ at each $x \in \{0, 12, 18, 23, 27, 30\}$, which we denote $g(x) - MAO(x)$ since they can be interpreted as the difference between the average proposer offer and the average responder MAO at x. See Section 3.3 for the full specification. Standard errors in Columns 1 and 2 are clustered around individuals, with individual characteristics (gender, undergrad/postgrad, faculty, previous experience with experiments, Australia born) included as regressors. Column 3 (4) reports the results of the non-parametric Fligner-Policello (Kolmogorov-Smirnov) test of differences in proposer offers and responder MAOs at each value of x. The coefficients in Columns 3 and 4 are the raw differences in means. 55 proposers at each x. The number of responders is 39, 39, 37, 21, 13, and 15 at the six increasing values of x.

Table 12: Tests of selfish-proofness of power disparities in treatment $\mathcal{CU}$

Y = offer/MAO	**Regression**		**FP**	**KS**
	Model 1	Model 2		
$g_s(0) - g_o(0)$	0.27		-0.48	-0.48
	(0.89)		(0.80)	(0.99)
$g_s(12) - g_o(12)$	1.76		1.15	1.15
	(0.14)		(0.13)	(0.39)
$g_s(18) - g_o(18)$	-0.62		-1.09	-1.09
	(0.67)		(0.44)	(0.83)
$g_s(23) - g_o(23)$	-4.39^{***}		-4.61^{***}	-4.61^{***}
	(0.01)		(0.01)	(0.01)
$g_s(27) - g_o(27)$	-5.29^{***}		-5.54^{***}	-5.54^{***}
	(0.01)		(0.01)	(0.01)
$g_s(30) - g_o(30)$			-8.41	-8.41
			(†)	(†)
x		-0.26^{***}		
		(0.01)		
$\mathbb{I}_p$		1.99		
		(0.26)		
$x \cdot \mathbb{I}_p$		-0.23^{**}		
		(0.03)		
x dummies	Yes			

Notes. p-values in parentheses. *: p-value ≤ 0.1 **: p-value ≤ 0.05 ***: p-value ≤ 0.01; p-values strictly less than 0.01 are rounded off to 0.01. †: different by construction. Columns 1 and 2 report results of random effects regressions where the dependent variable is the proposer's offer. Column 1 provides estimates of the coefficients of $\mathbb{I}_x \cdot \mathbb{I}_{SRP}$ at each $x \in \{0, 12, 18, 23, 27, 30\}$, which we denote $g_s(x) - g_o(x)$ since they can be interpreted as the difference in the average offer between self-regarding and other-regarding proposers at x. See Section 3.3 for the full specification. Standard errors in Columns 1 and 2 are clustered around individuals, with individual characteristics (gender, undergrad/postgrad, faculty, previous experience with experiments, Australia born) included as regressors. Column 3 (4) reports the results of the non-parametric Fligner-Policello (Kolmogorov-Smirnov) test of differences in offers between self-regarding and other-regarding proposers at each x. The coefficients in Columns 3 and 4 are the raw differences in means. 21 other-regarding and 34 self-regarding proposers used for all estimates.

Table 13: Robustness check: Is x fair or unfair?

Combination	**Treatment $\mathcal{U}$**		**Treatment $\mathcal{CU}$**	
	p-values	Classification	p-values	Classification
x = 0				
(DP, SP)	(0.05†, 0.41)	**Fair**	(0.07†, 0.80)	**Fair**
(DP, SP_1)	(0.05†, 0.03†)	Fair	(0.05†, 0.67)	Fair
(DP, SP_2)	(0.05†, 0.15)	Fair	(0.05†, 0.74)	Fair
(DP_1, SP)	(0.53, 0.41)	Fair	(0.18, 0.80)	Fair
(DP_1, SP_1)	(0.53, 0.03†)	Fair	(0.18, 0.67)	Fair
(DP_1, SP_2)	(0.53, 0.15)	Fair	(0.18, 0.74)	Fair
x = 12				
(DP, SP)	(0.25, 0.73)	**Fair**	(0.04†, 0.13)	**Fair**
(DP, SP_1)	(0.25, 0.68)	Fair	(0.04†, 0.76)	Fair
(DP, SP_2)	(0.25, 0.18)	Fair	(0.04†, 0.84)	Fair
(DP_1, SP)	(0.92, 0.73)	Fair	(0.13, 0.13)	Fair
(DP_1, SP_1)	(0.92, 0.68)	Fair	(0.13, 0.76)	Fair
(DP_1, SP_2)	(0.92, 0.18)	Fair	(0.13, 0.84)	Fair
x = 23				
(DP, SP)	(0.01, 0.01)	**Unfair**	(0.01, 0.01)	**Unfair**
(DP, SP_1)	(0.01, 0.01)	Unfair	(0.01, 0.01)	Unfair
(DP, SP_2)	(0.01, 0.01)	Unfair	(0.01, 0.11)	**Fair**
(DP_1, SP)	(0.01, 0.01)	Unfair	(0.01, 0.01)	Unfair
(DP_1, SP_1)	(0.01, 0.01)	Unfair	(0.01, 0.01)	Unfair
(DP_1, SP_2)	(0.01, 0.01)	Unfair	(0.01, 0.11)	**Fair**
x = 27				
(DP, SP)	(0.01, 0.01)	**Unfair**	(0.01, 0.01)	**Unfair**
(DP, SP_1)	(0.01, 0.01)	Unfair	(0.01, 0.01)	Unfair
(DP, SP_2)	(0.01, 0.01)	Unfair	(0.01, 0.04)	Unfair
(DP_1, SP)	(0.01, 0.01)	Unfair	(0.01, 0.01)	Unfair
(DP_1, SP_1)	(0.01, 0.01)	Unfair	(0.01, 0.01)	Unfair
(DP_1, SP_2)	(0.01, 0.01)	Unfair	(0.01, 0.04)	Unfair
x = 30				
(DP, SP)	(0.01, –)	**Unfair**	(0.01, –)	**Unfair**
(DP, SP_1)	(0.01, –)	Unfair	(0.01, –)	Unfair
(DP, SP_2)	(0.01, 0.01)	Unfair	(0.01, 0.01)	Unfair
(DP_1, SP)	(0.01, –)	Unfair	(0.01, –)	Unfair
(DP_1, SP_1)	(0.01, –)	Unfair	(0.01, –)	Unfair
(DP_1, SP_2)	(0.01, 0.01)	Unfair	(0.01, 0.01)	Unfair

Notes. The table reports *two-tailed* p-values under Fligner-Policello tests of dispute-proofness and selfish-proofness for the combination listed in the corresponding row. A power disparity is unfair if both dispute-proofness and selfish-proofness are rejected at the 10% significance level. See Table 3 in the text for the definitions of each combination. † indicates difference is significant but its sign is opposite to what is required for a violation of dispute-proofness or selfish-proofness. At each x, the fairness classification under the baseline (DP, SP) combination and deviations from the baseline classification are highlighted in bold. Some entries for selfish-proofness at $x = 30$ are vacant because of the underlying strategy used to identify self-regarding and other-regarding proposers.