PREFERENCES FOR POWER

ELENA S. PIKULINA AND CHLOE TERGIMAN

ABSTRACT. Power—the ability to determine the outcomes of others—usually comes with various benefits: higher compensation, public recognition, etc. We develop a new game, the Power Game, to demonstrate that a substantial fraction of individuals enjoy the intrinsic value of power: they accept lower payoffs in exchange for power over others, without any benefits to themselves. These preferences exist independently of other components of decision rights, cannot be explained by social preferences and are not driven by mistakes, confusion or signaling intentions. We further show that valuation of power (i) is higher when individuals *directly* determine outcomes of others; (ii) depends on how much discretion one has over those outcomes; (iii) is tied to relationships between individuals; and (iv) is likely domain dependent. We establish that ignoring preferences for power may have large welfare implications and, consequently, should be included in the study of political systems and labor contracts.

Date: **First Version: May 2017. Current version: August 2019**.

Key words and phrases. Preferences for power, private benefits of control, social preferences, other-regarding preferences, laboratory experiment

JEL: C91, D01, D03, M21.

Pikulina: elena.pikulina@sauder.ubc.ca, Sauder School of Business, University of British Columbia. Tergiman: cjt16@psu.edu, Smeal College of Business, Pennsylvania State University.

We thank Nageeb Ali, Gary Bolton, Gary Charness, Emel Feliz-Ozbay, Zack Grossman, Yoram Halevy, Paul J. Healy, Holger Herz, Elena Katok, Tony Kwasnica, Yusufcan Masatlioglu, Ryan Oprea, Erkut Ozbay, Carolin Pflueger, Doron Ravid, Ran Shorrer, Ron Siegel, Ennio Stacchetti, Severine Toussaert, Neslihan Uler, Emanuel Vespa, Marie-Claire Villeval, and Sevgi Yuksel for lively discussions and valuable feedback. We also benefited from comments from seminar participants at Bocconi University, The Sauder School of Business at UBC, Simon Fraser University, Pittsburgh University, The Technion (Haifa, Israel), GATE-LSE (Lyon, France), University of Zürich, University of Maryland, The Pennsylvania State University, The Naveen Jindal School of Business at UT Dallas, Chapman University, University of Utah, University of Toronto as well as participants at the following conferences: the Southwest Experimental and Behavioral Economics Workshop at UCSB, the Bay Area Behavioral and Experimental Economics Workshop at Santa Clara University, ESA San Diego, the TIBER Conference, the CESS 15th Anniversary Conference at NYU, SAET (Ischia, Italy). Pikulina and Tergiman are also very grateful for generous funding from the Social Sciences and Humanities Research Council of Canada (SSHRC).

> "When a moderate degree of comfort is assured, both individuals and communities will pursue power rather than wealth: they may seek wealth as a means to power, or they may forgo an increase of wealth in order to secure an increase of power, but in the former case as in the latter their fundamental motive is not economic."
>
> — Bertrand Russell, Power

Rational economic agents with standard preferences are interested in controlling the fates of others only as long as such power grants them immediate or prospective material benefits, for example, increases their payoffs, expands their choice set, or decreases risk and uncertainty. In this paper, we develop a new game, the "Power Game," to identify the intrinsic value of power by measuring how much people are willing to pay to increase it.

As it relates to the interaction between people, the Oxford Dictionary defines power as "the capacity or ability to direct or influence the behavior of others."[1] In social psychology, power definitions could be summarized as the asymmetric control over valued resources that allows individuals the ability to control the outcomes, experiences, or behaviors of others.[2] For example, Keltner, Gruenfeld and Anderson (2003) define power as the capacity to modify others' states by providing or withholding resources or administering punishments. Most importantly, the literature emphasizes that an individual's power should be characterized not in absolute terms but as falling on a continuum *relative* to the amount of potential influence as well as relative to the power of others (Fiske and Dépret (1996), Keltner, Gruenfeld and Anderson (2003)).

In this paper we study an important aspect of power: the ability to determine someone else's pay. In the workplace, relationships in which one party has the power to decide on another party's financial outcome are ubiquitous. In a principal-agent context for example, a principal's power often comes with the ability to decide on someone else's pay. Managers (or internal committees as in academia) regularly have the power to financially impact others via promotions, non-promotions, demotions, merit increases, etc. Moreover, managerial power is often relative in a sense that managers have more power when they can directly determine their employees' fines and bonuses and less power when their employees' entire compensation is pre-determined. Since individuals who value power are more likely to seek and attain positions of authority than others, the existence of such preferences has implications for the design of compensation contracts, promotion decisions and political systems.

Power, control, and autonomy all characterize capacity to make decisions. However, they have different connotations and illustrate different aspects of decision-making processes. For example, the decision-making capacity of a business owner includes *power*, the ability to determine outcomes of her employees, *control*, the ability to determine her own outcomes, and

[1] See Oxford Dictionary https://en.oxforddictionaries.com/definition/power.

[2] See, for example, Emerson (1962), Fiske (1993), Fiske and Dépret (1996), Keltner, Gruenfeld and Anderson (2003), Magee and Galinsky (2008), Tost (2015).

autonomy, the enjoyment of non-interference in her affairs by others.[3] Therefore, it is important to separate power preferences from those for autonomy and control. Indeed, control and autonomy are "inward-looking" and directed at the decision-maker herself, whereas power is "outward-looking" and by definition directed at others. Therefore, how individuals who enjoy power choose to exercise it has important implications for the interactions and relations between individuals. The Power Game allows us to focus solely on power, independently of control and autonomy, and is the first to identify intrinsic preferences for power, independent of any material benefits. In addition, it is designed to allow us to examine how individuals with such preferences choose to exercise power once in charge.

There are two main challenges in measuring individuals' preferences for power—the ability to determine payoffs of others—and in estimating their willingness to pay for it. The first challenge is that people may have social preferences and therefore choose to decide on payoffs of others because they put a non-zero weight on those payoffs. In particular, someone may not enjoy choosing the payoff of someone else, but may enjoy the resulting distribution.[4] The second challenge is to offer a set of choices that involve a meaningful trade-off between an individual's payoff and the amount of power she has over the payoffs of others. The Power Game meets both challenges. It has two parts. In Part I, there is an explicit trade-off between a player's payoff and her ability to determine payoffs of another player, which an individual with power preferences can exploit. In contrast, Part II does not offer such a trade-off but instead controls for the payoff values from Part I. By offering different power-payoff trade-offs in Parts I and II and controlling for the payoff values in Part II, our design allows us to isolate choices due to power preferences from those potentially explained by social preferences.

In the Power Game, there are two types of players, A and B, who are matched in pairs. Only type A players make decisions, and these decisions determine the payoffs of both A and B players. In Part I, A chooses between two options. In the first option, both A and B receive E_A, hence, the resulting allocation is (E_A, E_A). Player A's second option is to receive a different payoff, $E_A - p$, and obtain the right to choose a specific payoff for B in the $[0, E_B]$ interval. In other words, if A pays price p, then $(E_A - p, x_B^*)$ is the resulting allocation, where x_B^* is what A chose for B in the $[0, E_B]$ interval. Because Part I has several rounds and p varies from round to round, we can determine an individual's willingness to pay for the right to determine B's payoff.[5]

In Part II, player A makes choices between two payoff pairs that determine payoffs for herself and B. Each round in Part I has a corresponding round in Part II. If in Part I A paid price p, then in the corresponding round of Part II she has to choose between $(E_A - p, x_B^*)$ and

[3]Control and autonomy are not synonymous. Consider, for example, the case where an individual's payoff is determined randomly. In this case, she has no control but does have autonomy.

[4]Alternatively, one may enjoy being seen as a kind person. Our experimental implementation distinguishes preferences for power from such signaling motives (see Section 1 for details).

[5]The price p can be positive, i.e. A incurs a cost in order to choose for B. Alternatively, p can be negative, i.e. A is compensated for choosing for B. In our experimental implementation, we use the following parameter values: $E_A = \$12.30$, $E_B = \$16.30$, and the price p varies from $-\$0.25$ to \$2 in increments of 25 cents.

(E_A, E_A), where x_B^* is her choice for B in Part I. In other words, she has the choice between the allocation she actually chose in Part I, $(E_A - p, x_B^*)$, and (E_A, E_A), the allocation she could have chosen if she didn't pay. If in Part I A did not pay p, then in the corresponding round of Part II she has to choose between (E_A, E_A) and $(E_A - p, E_A + 2p)$, the allocation she actually chose and a more efficient one that she could have chosen (see Section 1.3 for details).

While A has power over B's payoff in both Parts of the Power Game, she faces different trade-offs between the amount of power and her own payoff in Parts I and II. It is only in Part I that A can acquire more power. Indeed, in Part I, when A pays p, she obtains the right to choose B's payoff precisely, and can choose any payoff she pleases within the $[0, E_B]$ interval. If, on the other hand, A does not like power, then when p is negative, she can forgo a payoff increase in order to avoid choosing for B. Thus, there is an explicit trade-off between A's payoff and A's power over B, which an individual with power preferences can exploit. In contrast, Part II does not offer such a trade-off but instead controls for x_B^*, A's choice for B from Part I. When A gives up p in Part II and chooses the payoff pair with the lower payoff for herself, it does not change her power over B but simply implies a different, fixed, payoff for B. Having different power-payoff trade-offs for A in Parts I and II and controlling for her choices for B allows us to determine why she paid in Part I. Did she pay because she desired a specific outcome $(E_A - p, x_B^*)$? Or did she pay because she enjoyed the power of choosing B's payoff in $[0, E_B]$ but in fact attached little importance to her actual choice of x_B^*?

By comparing how much subjects are willing to pay in Parts I and II of the Power Game, we are able to classify their preferences. While players with standard preferences never pay positively in both Parts, players who value power or have social preferences pay non-zero prices in Part I. Players with power and social preferences, however, behave differently in Part II. If A's choices in Part I are the result of her social preferences and she does not place any value on the process by which final allocations are attained, then in Part II she should still prefer $(E_A - p, x_B^*)$, the allocation she implemented in Part I. In other words, player A should be willing to pay price p to implement her desired allocation irrespective of whether she picks B's payoff herself as in Part I, or whether the exact same payoff is exogenously given as in Part II. If, in contrast, in Part I, player A pays only to increase her power over B, then in Part II she should prefer (E_A, E_A), since paying in Part II does not lead to any additional power but simply lowers her payoff. Thus, if a player reverses her choices in Part II and chooses (E_A, E_A) instead of $(E_A - p, x_B^*)$, then she must have preferences for power. In other words, subjects who have preferences for power enjoy the *process* of choosing payoffs of others, without receiving additional utility from the actual payoff itself.

Our main finding is that about 28% of subjects have preferences for power without social preferences. We call them Power+ subjects. These subjects are willing to pay over 10 percent of their potential payoff to be able to choose payoffs for B in Part I, but they are willing to pay nothing to implement the *same* allocations in Part II, when additional power is not attainable. Subjects who have social preferences in any capacity (i.e. with or without power preferences) represent about 19% of our subjects. In total, subjects who have preferences for power in any capacity (i.e. with or without social preferences) constitute about 36% of our subjects. Finally, about 47% of subjects have standard preferences.

We then provide evidence that our Power-Game-based preference classification indeed captures differences in preferences across subjects. Since our classification depends only on the difference in subjects' willingnesses to pay across Parts I and II, we can use it to predict other behaviors of subjects. We show that subjects we classify as having social preferences, regardless of their attitude towards power, are consistent in the amounts they give to B: 93.8% of them always give the maximum allowable amount, E_B. In contrast, Power+ subjects exhibit much more variation in their giving behavior both within and across subjects: they choose amounts that span almost the entire choice space, that is, the $[0, E_B]$ interval, implying that social welfare is likely to decrease when power-hungry individuals are the ones allocating resources. In addition, we show that these preference classes predict subjects' decisions in tasks that are unrelated to Part I. More specifically, when additional power is not attainable, subjects with power preferences behave much like subjects with standard preferences, that is they maximize their own payoff, while those with social preferences do not.

Note that Part II of the Power Game is designed to control for outcome-based social preferences as a reason to pay in Part I. Our experiment also takes care to minimize the role of intentions-based social preferences. We do so by ensuring that subjects cannot attribute their final payoffs to their own actions or to the actions of others, which has shown to largely weaken the reciprocal response between individuals, a central tenet of intentions-based social preferences. Indeed, our data support the notion that the behavior of Power+ subjects is unlikely to be explained by such models of behavior. For example, the pattern (or lack thereof) in Power+ subjects' giving behavior is at odds with models of intentions-based social preferences, self-signaling or interdependent preferences, which posit that subjects often want to appear nice (either to themselves or to others) or want to avoid feeling guilty. Contrary to these models, in Part I, Power+ subjects do not consistently make choices that put them in a good light since they infrequently give the maximum allowable amount E_B. In Part II, Power+ subjects revert *all* their decisions, i.e. both the ones in which $x_B^* < E_A$ and those in which $x_B^* > E_A$.

To further understand what drives individuals' valuation of power, we conduct four additional treatments. Three of those are "modified" Power Games, the fourth is a vignette study. In the first modified Power Game, we change the maximum allowable amount A can give to B, such that $E_B = E_A$. In other words, when A pays she chooses x_B from the $[0, E_A]$ interval and when she does not pay, she implements the (E_A, E_A) allocation. That is, we reduce the size of A's choice set and remove kind/efficient choices. The fraction of Power+ subjects in this treatment is statistically no different than in our main treatment. However, subjects' willingness to pay for power is reduced. Thus, the magnitude of the utility derived from power depends on the nature of choices players can make.

In the second modified Power Game, subjects are paired with a charity instead of with another player. Here we observe only a negligible fraction of Power+ subjects. Further, their willingness to pay also decreases sharply. These results show that Power+ subjects are not driven by the "lure of choice" (Bown, Read and Summers (2003)), since the choice space here and in the main treatment are the same. This also suggests that the distance between a decision maker and the "other" as well as the impact on the "other" may matter for the perception and value of power.

In the third modified Power Game, A can pay for the right to *influence* as opposed to *determine* B's payoff. More specifically, if A pays, a computer randomly chooses B's payoff from the $[0, E_B]$ interval. In this treatment, Power+ subjects also virtually disappear, in line with the work by Ferreira, Hanaki and Tarroux (2017) and Neri and Rommeswinkel (2017), who found that subjects are not willing to pay for the ability to affect payoffs of others in a probabilistic way. This result shows that individuals attach more value to their ability to directly *determine* outcomes of others as opposed to *influence* those outcomes in a probabilistic way.

Finally, we also conduct a vignette study to illustrate individuals' perception of power by third parties in different real-life situations. In these vignettes, subjects are exposed to 5 hypothetical everyday workplace and university-life scenarios. We show that in the eyes of subjects power is increased when an individual can choose others' payments from an interval. This result provide further support for the assumption that subjects can indeed increase their power by paying p in the Power Game. Looking at the difference in subjects' responses in our workplace and university-life scenarios, we are also find that whether value of power increases with choice set size likely depends on the domain.

The study addresses an important issue in the recent experimental literature on preferences for decision rights. While prior studies have shown that individual preferences for decision rights exist, they have not been able to disentangle their various components. For example, Owens, Grossman and Fackler (2014) find that when asked whether to bet on their own performance or on their partners' performance in a quiz, people prefer to bet on themselves. Although the "illusion of control" may explain some of the individuals' choices to retain decision rights (e.g. Sloof and von Siemens (2017)), it remains unclear whether people prefer to retain decision rights because they like having control over their own payoffs or because they are averse to losing their autonomy to others. Similarly, Fehr, Herz and Wilkening (2013) find that principals do not delegate decision rights to agents often enough in games where delegation results in higher monetary payoffs for both parties. While regret aversion may account for a portion of the retained decision rights, Bartling, Fehr and Herz (2014) show that under-delegation is also driven by individuals assigning a positive value to decision rights *per se*. They however acknowledge that their "design does not allow disentangling whether a possible positive intrinsic value of decision rights stems from the desire to be able to affect someone else's payoffs or from the aversion to be affected by some else's decision" (p.2022). Our paper is the first one to provide evidence that individuals value power *per se*, i.e. their ability to determine payoffs of others. Given the behavior of power-hungry subjects, it is important to identify those since having them in top positions can have dramatic implications on welfare of others.

In addition, our findings contribute to the corporate finance and delegation literatures that consider the private benefits of decision-making as one of the main frictions in the principal-agent problem and in optimal organizational design (e.g., Jensen and Meckling (1976), Grossman and Hart (1986), Aghion and Bolton (1992), Hart and Moore (2005), Dessein and Holden (2019)). The theoretical literature has pointed out the possible non-pecuniary nature of private benefits. Hart and Moore (1995), for example, motivate their theory by claiming that "among other things, managers have goals, such as the pursuit of power" (p. 568). By their very nature, non-pecuniary private benefits are difficult to observe and even more difficult to quantify in a reliable way. Instead, the empirical literature has concentrated on measuring pecuniary private benefits by estimating the value of perquisites enjoyed by top executives (Demsetz and Lehn (1985), Barclay and Holderness (1989), Dyck and Zingales (2004), Dahya, Dimitrov and McConnell (2008), Doidge et al. (2009)). For example, Dyck and Zingales (2004) find substantial evidence that good institutions and corporate governance can significantly curb the amount of monetary private benefits enjoyed by controlling shareholders. However, our results call into question whether even the best institutions would be able to eliminate non-pecuniary private benefit frictions in the presence of power-hungry agents.

Our results are also related to the literature on procedures versus outcomes. In strategic games, when evaluating decisions of others, individuals may base their assessments not only on outcomes but also on the procedures that lead to those outcomes. Indeed, past work has shown that including a third party in the decision-making process, changing the distance between a decision-maker and a recipient, varying the possibility of retribution and modifying the interpretation of motives leads individuals to evaluate outcomes differently. This is the case, for example, in Fershtman and Gneezy (2001), Coffman (2011), Bartling and Fischbacher (2011), and Orhun (2018). In our paper, we show that a large fraction of individuals care about procedures when it comes to how they *themselves* reach decisions concerning *others*, as opposed to how someone else acts towards them or others. This is the case even in the absence of strategic interactions, any possibility of retribution and in situations where beliefs regarding others' subsequent actions are irrelevant.

Finally, our findings have important methodological implications for inferring social preferences from individual choices. For example, Lazear, Malmendier and Weber (2012) show that sharing in dictator games decreases when individuals are allowed to opt out. Furthermore, Zizzo and Oswald (2001), Abbink and Sadrieh (2009), and Charness, Masclet and Villeval (2014) show that when people can choose by exactly how much to decrease the payoffs of others, many of them are willing to sacrifice their own payoffs in order to "burn" other people's money. However, our study demonstrates that a large fraction of the population has preferences for power, and individuals with such preferences may appear spiteful if their only option is to decrease the payoff of others even though they do not attach any value to those payoffs *per se*.

Our study reconciles results from these above papers with those studies that have shown that when people can only pick between two fixed options, where one of the options gives them less money but also destroys the payoffs of their partners, they behave in a much less malicious way (Charness and Rabin (2002), Chen and Li (2009)).

The remainder of the paper is organized as follows. We detail the Power Game and its experimental implementation in Section 1. In Section 2 we derive a set of theoretical predictions for the subjects' behavior in the Power Game. Section 3 presents the main experimental results. In Section 4, using four additional experiments, we investigate potential mechanisms underlying preference for power. Section 5 discusses and refutes potential alternative explanations of our results. Section 6 concludes.

1. Experimental Design: The Power Game

1.1. **The Power Game.** We develop a new game, the "Power Game" and describe it here. The game has two parts. At the beginning of Part I, players are randomly assigned a type, either A or B, with equal number of type As and type Bs. Types are fixed throughout the entire game and only type A players make decisions.

Part I: Part I comprises N rounds. In each round, player A is randomly matched with player B. In round j, a price p_j is revealed to A who must then decide whether to pay it or not.

- If player A pays p_j, then the payoffs for the players are $(E_A - p_j, x^*_{Bj})$, where x^*_{Bj} is what A chooses for B in the $[0, E_B]$ interval.
- If player A does not pay p_j, then the payoffs for the players are (E_A, E_A).

The values of E_A and E_B are known in advance and fixed throughout all the rounds. In each round, for each A, the price p_j is randomly drawn from a discrete set $\mathcal{P}$, of size N, without replacement, and revealed to players before they make a decision on whether to pay it or not.

Part II: Part II lasts for M rounds where $M \geq N$. In each round, player A decides between two payoff pairs: (x_A, x_B) and (x'_A, x'_B). N of the M rounds correspond to the N Part I rounds. These rounds are player-specific as they depend on a player's decisions in Part I of the Power Game. More specifically, for each $p_j \in \mathcal{P}$:

- If in round j of Part I player A paid p_j, then in the corresponding round of Part II, she decides between the following payoff pairs: $(E_A - p_j, x^*_{Bj})$ and (E_A, E_A), where x^*_{Bj} is the payoff she chose for player B in round j of Part I.
- If in round j of Part I player A did not pay p_j, then in the corresponding round of Part II, she chooses between (E_A, E_A) and $(E_A - p_j, E_A + 2p_j)$.

Whether or not a player paid p_j in round j of Part I, one of the payoff pairs she faces in the corresponding round of Part II is the pair she actually chose in Part I: $(E_A - p_j, x^*_{Bj})$ for players who paid and (E_A, E_A) for those who did not. The other payoff pair she faces is one she could have chosen in round j of Part I but rejected: (E_A, E_A) if the player paid p_j and $(E_A - p_j, E_A + 2p_j)$ if she did not pay (see Section 1.3 for more details). Importantly, for each p_j a player encountered in Part I, in Part II she faces a choice between two payoff pairs, one of which is *identical in payoff distribution* to the pair that she actually selected in Part I, and the other is a pair she rejected.

Note that player A has power over B's payoff in both Parts of the Power Game. However, she faces different trade-offs between power and her own payoff in Parts I and II. If A pays p_j in Part I, she increases her power over B's payoff since she can select any number in the $[0, E_B]$ interval, including E_A. If she doesn't pay, then she effectively chooses the only payoff option available for B, i.e. E_A. In Part II, whether player A pays p_j or not, she still chooses only from a single payoff option for B, either x_B^* or E_A. In other words, when paying in Part II, player A does not acquire more power but instead obtains a different pre-specified fixed payoff for B.

The payoff pairs in the remaining $M - N$ rounds in Part II are chosen independently of Part I and correspond to other choices that may be of a separate interest to the researcher.

1.2. **Experimental implementation.** All our Power Game experimental sessions were conducted in February 2018 at the Laboratory for Economic Management and Auctions (LEMA) at the Pennsylvania State University using z-Tree software (Fischbacher (2007)). Subjects were recruited from the general undergraduate population and each subject participated in one session only. We conducted 16 sessions for a total of 288 subjects. Each session lasted at most 1 hour and on average participants earned $18 (the median was $19.30), including the show-up fee of $7.

Our experimental design consisted of four tasks. The first task was a simple lottery task. Subjects then participated in Part I of the Power game. Part II of the Power Game took place directly afterwards. Finally, the last task subjects faced was a repeat of Part I of the Power Game; heretofore, we refer to this last task as Part I*. Instructions for each task were handed out and read out loud after the previous task had been completed. Subjects were told that only one of their decisions, randomly chosen, would count for their payment. They were also told that at the end of the experiment, the only information that they would receive would be their total earnings. Before leaving the lab, subjects filled out a questionnaire where we asked them what motivated their choices, as well as demographic and education information. The full set of instructions is in Appendix A, examples of the game interface are in Appendix B, and the final questionnaire is in Appendix C.

The Lottery task: The Lottery task consisted of five rounds. In each round, subjects decided between receiving a fixed amount and a random uniform draw from the $[\$0, \$16.30]$ interval, at five cent increments. The fixed amounts were drawn without replacement from $\{\$0, \$3.10, \$6.60, \$12.30, \$16.30\}$ and subjects faced them in random order. Which option appeared on the left or the right of the screen was randomly and independently determined for each subject. In addition, the fixed option amounts were listed explicitly in the instructions so subjects were aware of the specific choice problems they and others would be facing over course of the Lottery task.[6]

[6]Due to a technical issue, we were not able to collect these data for one of the sessions, which affected 22 subjects.

The Power Game: After the Lottery task had been completed but before Part I of the Power Game, subjects were randomly assigned a type: A or B. Subjects were told that throughout the rest of the experiment only type A players' decisions would matter for payment and that types would remain fixed. Subjects however were not told what type they were, but asked to make decisions as if they were type A players. If their *true* type was B, none of their decisions would matter for payment. If their *true* type was A, then one of their decisions, randomly selected, would matter. Thus, regardless of one's true type, it was in one's best interest to make decisions as if one were a type A player. True types were never revealed to the subjects.

In each round, each A player was randomly matched with a B player. Subjects moved from one round to the next when all subjects had completed the previous round. Before starting Part I of the Power Game, subjects were shown three screens (see Appendix B). In the first of those three screens, they were shown what a first stage screen of Part I would look like. They were then shown the screens that paying and not paying would lead to. Thus, they could familiarize themselves with the interface and satisfy any curiosity regarding what paying or not would lead to in terms of screen display.

Instructions for Part II were handed out and read out loud after Part I was completed. Thus, our subjects were not aware of the contents of Part II when they were making their Part I decisions. After the end of Part II, we handed out instructions for the final task of the experiment, Part I*. Those instructions were identical to those subjects received the first time they played Part I of the Power Game, save for an opening sentence telling them the task would be the same and that these new instructions served to remind them of the task.

Parameter values in Part I: We used the following parameter values in Part I: $E_A = \$12.30$ and $E_B = \$16.30$. The set $\mathcal{P}$ contained 10 distinct prices ranging from $-\$0.25$ to \$2, in increments of 25 cents: $\mathcal{P} = \{-\$0.25, \$0, \$0.25, ..., \$1.75, \$2\}$. Thus, subjects played a set 10 rounds where prices were randomly and independently drawn for each subject in each round without replacement from $\mathcal{P}$, with the exception of the negative price of $-\$0.25$, which was drawn in round 10 for all subjects.

Subjects were not aware of the contents of $\mathcal{P}$, they were simply told that the price would vary from round to round. If A decided to pay, she would receive $\$12.30 - p$ as her payoff and she would obtain the right to choose the payoff for B, and could choose any number between \$0 and \$16.30 (in increments of 5 cents). If A did not pay, then both A and B would each receive \$12.30.

Round 11 of Part I: After all 10 prices in $\mathcal{P}$ had been drawn, subjects played an additional round, where they had the choice between the $(0, 0)$ and $(12.30, x_B)$ payoff pairs, where x_B is A's choice for B in the $[0, 16.30]$ interval. The expectation was that all subjects would choose the latter option and indeed all subjects did so. This round was included so that we could see what *all* subjects choose for B when their own payoff was 12.30, since at a price of 0, subjects can still choose not to pay.

Part I screens: The first screen subjects faced in each round clearly showed the two payoff pairs that subjects had to choose from (see Appendix B). Which option was on the left or on the right was randomly determined for each subject in each round. In each round of Part I of the Power Game, after deciding whether to pay or not, *all* subjects faced a second screen. If a subject paid p, she would then have to enter the amount she wished to give to B. If a subject did not pay p, she would have to enter between 1 and 5 characters of his/her choice (numbers, letters and special characters were all allowed).

Parameter values in Part II: Part II consisted of 22 rounds where subjects decided between two payoff pairs. Which payoff pair was presented on the left or on the right of the screen was randomly determined for each subject in each round. 10 rounds were subject-specific and 12 rounds were identical for all subjects. The order of rounds was random for each subject.

In Part II, the 10 subject-specific rounds depended on a particular subject's decisions over the first 10 rounds of Part I. Specifically, subjects decided between the payoff pair they chose in Part I and a pair that was available but rejected:

- If a subject paid p_j and chose x^*_{Bj} in round j of Part I, she had to choose between the following payoff pairs in the corresponding round in Part II: $(12.30 - p_j, x^*_{Bj})$ and $(12.30, 12.30)$.
- If a subject did not pay p_j in round j of Part I, she had to choose between the following payoff pairs in the corresponding round in Part II: $(12.30, 12.30)$ and $(12.30 - p_j, 12.30 + 2p_j)$.

The remaining 12 rounds were identical for each subject. In six of those rounds, the values for the payoffs pairs were inspired by Charness and Rabin (2002)[7] and re-scaled such that the order of magnitude for payoffs was similar to the values stemming from Part I, see decisions CR1-CR6 in Table 1. Other decision problems were chosen to be similar to some of the problems in Charness and Rabin (2002) but to allow for different trade-offs between the payoffs of players A and B, see decisions PT1-PT3 in Table 1. Decision problem PT4 was designed to check whether subjects understood that they were to act as type A players. Finally, problems PT5 and PT6 were chosen to serve as "sanity checks" in our analysis.

1.3. **Design choices.** A few elements of our design are worth elaborating upon.

- The Lottery task. The Lottery task was included to ensure that subjects would be unable to tell both their type and what task of the experiment was chosen to count for payment. Indeed, regardless of what task was chosen to count for payment, and regardless of what type a player was, that payment could have come from the Lottery task. This curtails intention-based motivations as subjects see that any payment B receives can be the result of his own decisions in the Lottery task and not necessarily the results of A actions (see also Section 5.1 for a detailed discussion).

[7]See two-person dictator games, Table 1, p. 829.

TABLE 1. Decision problems in 12 rounds of Part II.

Decision[a]	First Option[b]	Second Option
CR1	$(6.60, 6.60)$	$(6.60, 12.30)$
CR2	$(6.60, 6.60)$	$(6.20, 12.30)$
CR3	$(10.50, 5.30)$	$(8.80, 12.30)$
CR4	$(12.30, 3.50)$	$(10.50, 10.50)$
CR5	$(12.30, 0.00)$	$(6.15, 6.15)$
CR6	$(3.10, 12.30)$	$(0.00, 0.00)$
PT1	$(10.10, 5.20)$	$(9.10, 9.10)$
PT2	$(12.30, 5.10)$	$(10.10, 12.30)$
PT3	$(12.55, 12.80)$	$(12.30, 12.30)$
PT4	$(12.30, 9.60)$	$(9.60, 12.30)$
PT5	$(12.30, 7.80)$	$(7.80, 5.40)$
PT6	$(6.15, 6.15)$	$(0.00, 0.00)$

[a]These rounds were presented among the 22 rounds of Part II in random order for each subject.
[b]What option was presented on the left or on the right of the screen was randomly determined independently for each decision problem and for each subject.

- Making $E_A < E_B$. This choice allows us to explore the interaction between power and social preferences. Our results are robust to making $E_B = E_A$ (see Section 4.1).
- Payoffs are (E_A, E_A) when A does not pay p. This is done for several reasons. The (E_A, E_A) choice is a natural focal point as both players earn identical amounts. One payoff pair being symmetric further isolates us from concerns related to inequality aversion or spitefulness. Thus, choosing the $(E_A - p, x_B^*)$ option is then more likely to be a deliberate action as opposed to choosing (E_A, E_A), which is more fair or salient.
- $(12.30 - p_j, 12.30 + 2p_j)$ is the alternative payoff pair in the Part II rounds that correspond to the Part I rounds in which subjects did not pay p_j. The advantage is three-fold. First, there is more variability in what subjects see as opposed to having a fixed alternative amount for player B. Second, prior literature indicates that most people are generous towards others rather than mean and prefer efficient allocations to inefficient ones. Finally, the efficiency gain increases with price: the more the subjects give up, the higher the efficiency gain.[8]
- Having all subjects type something after the decision to pay or not. The purpose of this is three-fold. First, this ensures that no subject could guess whether a neighbor had chosen to pay or not since everyone had to type in the second stage. Restricting the number of characters to be between 1 and 5 ensures that whatever was typed could have indeed been a number between 0 and 16.30. Thus, the anonymity of subjects' choices was preserved. Then, we mitigate any experimenter-demand effect where subjects might

[8]The potential downside of such a choice, however, is that we may mis-identify subjects who are both averse to power and have competitive social preferences (such subjects derive additional utility when their payoffs are higher than payoffs of others, see Charness and Rabin (2002)). This issue is not one that appears to matter in practice. We return to this point and provide additional details in Section 3.2.2.

pay in Part I because it is the only option with a subsequent action (de Quidt, Vesterlund and Wilson (2018)). Finally, it minimizes decisions to pay that would be due to boredom.

- Before Part I starts, subjects are told that throughout the remainder of the experiment, types are fixed and only A players make decisions that matter for payment. These design elements minimize the possibility that subjects' decisions are motivated by their belief that those decisions may be rewarded or used against them in some way by other subjects in other rounds/tasks of the experiment. To further ensure this, at the very start of the experiment, subjects are told that no choice they make in any given round/task can increase or decrease their potential payoff in any other round or task of the experiment.
- Subjects are not told what type they are. This design feature allows us to collect decisions from all our subjects since they all behave as if they were type A players, as opposed to revealing types and only collecting data from half of the subjects in each session. Our instructions carefully describe this design element and subjects are emphatically told that they should act as type A players, since if their true type were B none of their decisions would matter. In one of the rounds of Part II we directly test whether subjects understood their roles and find strong evidence that they did. In that specific round, all subjects are faced with a choice between two payoff pairs, $(12.30, 9.60)$ and $(9.60, 12.30)$ and 98% of our subjects choose $(12.30, 9.60)$. Had there been any doubt on who to make decisions for, the fraction choosing the latter would have been higher.[9]
- Unordered prices: p is randomly drawn from $\mathcal{P}$. In some experimental designs, the experimenter restricts the choices of subjects so that they appear rational and "well-behaved," e.g., such that all subjects have cutoff strategies. In our context this would mean imposing that as soon as a subject does not pay for some price, we force that the rest of her decisions be "not pay" for any price greater than that first price. Another way to "encourage" well-behaved choices is to offer an ordered list of prices. We however let price p be randomly drawn from $\mathcal{P}$ and ask subjects to make decisions for all prices in $\mathcal{P}$, regardless of past behavior. We do so for two reasons. First, we are able to identify the subset of subjects who *are* well-behaved and conduct several analyses: using those subjects only and using the entire sample. We can evaluate whether our results depend on the kind of subjects we are considering.Second, random price ordering ensures that our results are not driven by order effects.
- Having a negative price and leaving it for the tenth round. We incorporate the negative price in order to identify subjects who may be averse to power. Indeed, power-averse subjects would need to be compensated in order to choose for B. However, most subjects are not familiar with the concept of negative prices and keeping it for the last round

[9]Regarding how this implementation might impact subjects' choices to pay and what to give, we refer the reader to Brandts and Charness (2011). In a comprehensive review of the strategy method (Selten (1967)) they find that while the strategy method may intensify pro-social behavior, it does not fundamentally alter subjects' preferences.

ensures that prior choices in Part I were not affected by the presence of this negative price.

- Subjects play Part I twice. This feature was included to evaluate the robustness of our results and test whether behavior is consistent across Parts I and I*. Further, having another task after Part II ensures that subjects' behavior in that part is not affected by it being the final task in the experiment. In addition, this allows us to rule out explanations based on subjects changing preferences over time.[10]

2. Theoretical Framework

In this section, we present a set of predictions for Parts I and II of the Power Game for individuals in different preference classes. We think of individual preferences as varying along two dimensions. The first is whether an individual non-trivially incorporates other players' payoffs in her utility function. The second is whether she derives utility from having power over the payoffs of others. Thus, we consider the following four types of preferences: standard selfish preferences, power preferences, social preferences, and, since power and social preferences are not mutually exclusive, preferences that have both social and power components. The assumptions behind, and formal derivations of, these predictions are in Appendix D.

Experimental predictions and empirical identification of preference classes. Following our theory, Table 2 summarizes the correspondence between paying behavior across Parts I and II of the Power Game and preference classes. We expect subjects with standard preferences to never pay positive prices in either Part I or II. Subjects with power preferences never pay in Part II, but pay in Part I, positively if they enjoy power and only negatively if they are averse to it. Subjects with social preferences pay strictly positive prices in order to choose the payoffs for others in Part I and are willing to pay those same prices in Part II. Subjects who have social preferences and value power are willing to pay higher prices in Part I than in Part II, but still pay positively in Part II. If subjects have social preferences but dislike power, they pay more in Part II than in Part I. Note that in all cases subjects' willingness to pay depends on the strength of their power and social preferences. Finally, we cannot positively assign preference classes to subjects who pay at some prices in Part I but never pay in Part II.

There are three caveats to our identification of preferences with a power component. Before we describe them in turn, we point out that they may bias our results in one direction only: the under-identification of the overall fraction of subjects with a power component to their preferences.

The first caveat is related to the experimental implementation of the Power Game. Since in the experiment subjects face a menu of prices in increments of 25 cents, we are unable to observe subjects' true willingnesses to pay in Parts I and II but only their lower bounds in increments of 25 cents. In particular, subjects whose willingness to pay is strictly less than 25 cents in absolute value are treated as if their true willingness to pay is \$0. In this case,

[10] Brosig-Koch, Riechmann and Weimann (2017) show that subjects may become less pro-social over time.

TABLE 2. Empirical identification of preference classes.

Preference class	$\bar{p}_{\text{I}}$	$\bar{p}_{\text{II}}$
Standard[a]	0	0
Power +	$\bar{p}_{\text{I}} > 0$	0
Power −	$\bar{p}_{\text{I}} < 0$	0
Social Preferences	$\bar{p}_{\text{I}} > 0$	$\bar{p}_{\text{I}} = \bar{p}_{\text{II}}$
Social Preferences & Power +	$\bar{p}_{\text{I}} > 0$	$\bar{p}_{\text{I}} > \bar{p}_{\text{II}}$, $\bar{p}_{\text{II}} > 0$
Social Preferences & Power −	Any	$\bar{p}_{\text{I}} < \bar{p}_{\text{II}}$, $\bar{p}_{\text{II}} > 0$
Unclassified	Any	$\bar{p}_{\text{II}} < 0$

[a]Subjects with $\bar{p}_{\text{I}} = \bar{p}_{\text{II}} = 0$ might instead have strong preferences for equality. The behavior of 2.5% of our subjects might be consistent with such preferences. We choose to identify subjects for who $\bar{p}_{\text{I}} = \bar{p}_{\text{II}} = 0$ as having standard preferences with the caveat that a minority might have very strong preferences for equality.

we may underestimate the fraction of people with power preferences (or social preferences for that matter) as they may instead appear to us as having standard preferences. More generally, a subject should have strong enough relative preferences for us to identify those.

Next, subjects who have preferences for equality and enjoy power may pay in Part I and not in Part II, exactly like subjects who have power preferences only. We choose to categorize subjects who only pay positively in Part I as having power preferences only, with the caveat that a small fraction may in fact also have social preferences.[11]

Finally, subjects who have certain social preferences and dislike power may not necessarily pay more in Part II than in Part I, because in Part II they face an alternative with an efficiency trait, $(12.30 - p, 12.30 + 2p)$. In particular, subjects with competitive preferences or weak preferences for efficiency might find the $(12.30 - p, 12.30 + 2p)$ allocation unattractive compared to the $(12.30, 12.30)$ one. Such subjects would appear to us as belonging to the Social, Standard or Power- preference classes instead of the Social&Power- class.[12] Thus, we possibly mis-classify subjects with weak efficiency preferences and aversion to power into other preference classes.

3. Experimental Results

3.1. **Preferences for power: the aggregate level.**

3.1.1. *Demand for choosing payoffs of others in Part I.* We begin by exploring the relationship between prices and subjects' decisions to pay for the right to choose precise payoffs of others in Part I. Figure 1 presents the fraction of subjects who agree to pay in Part I for each given price. Three features deserve emphasis. First, the fraction of subjects who pay to choose

[11]Empirically, the fraction of such subjects is at most 0.6% of our sample, since this is the proportion of subjects who choose \$12.30 as B's payoff when their own payoff is \$12.30.

[12]We identify at most 2 subjects (1.2% of our sample) in the Standard class who might have competitive preferences and dislike power; no subjects in the Social class have competitive preferences, and there is only one subject in the Power- class. We acknowledge that Weak Efficiency&Power- preferences are much harder to pin down with our design. However, even if we used an alternative of $(12.30 - p, 16.30)$, for subjects with weak efficiency preferences \$16.30 could still not be high enough.

others' payoffs is decreasing in price, i.e., the demand function is downward sloping. Then, at the negative price of $-\$0.25$, 272 out of 288 subjects agree to increase their own payoff in exchange for choosing the payoff for B. Finally, at the price of \$0, about 81% of our subjects prefer to choose x_B. Thus, our data indicate that subjects understand there is a real cost to choosing B's payoff. The existence of an aggregate downward-sloping demand function shows that the demand for choosing payoffs of others is well-behaved at the aggregate level.

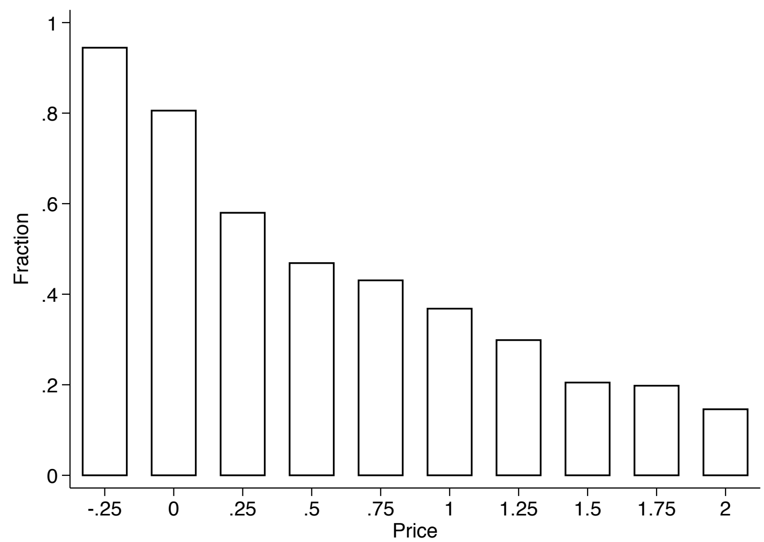

FIGURE 1. Fraction of subjects who pay to choose payoffs of others in Part I, by price in USD.

In Table 3 we present statistics regarding subjects' decisions to pay. For each subject, we calculate how many times he or she pays. The majority of subjects, about 64%, pay a positive price at least once. Further, 53.1% of our subjects pay positively at least twice and about 8.3% pay at all prices, including the price of \$2, which represents a loss of over 16% of what they could have earned had they not paid. In fact, the median number of strictly positive payments is 2, and the mean is 2.7.

TABLE 3. Fraction of subjects paying to choose payoffs of others in Part I.

	Subjects	Percentage
Never paid, including at price of $-\$0.25$	8	2.8%
Paid only at prices of \$0 and/or $-\$0.25$	103	35.8%
Paid a positive price at least once	185	64.2%
Paid a positive price at least twice	153	53.1%
Paid a positive price at least three times	130	45.1%
Paid at all prices, including \$0 and $-\$0.25$	24	8.3%
All subjects	288	100.0%

3.1.2. *Allocations chosen in Part I.* In Part I of the Power Game, our 288 subjects make 2,880 different decisions over the courses of the 10 prices they face. In 1,600 (or 55.6%) of the cases, subjects don't pay, i.e. choose the $(12.30, 12.30)$ allocation. Here we concentrate on the remaining 1,280 (or 44.4%) of the allocations, when subjects pay p, i.e. choose the $(12.30 - p, x_B^*)$ allocation. We depict those allocations on the $x_A x_B$ plane in Figure 2.

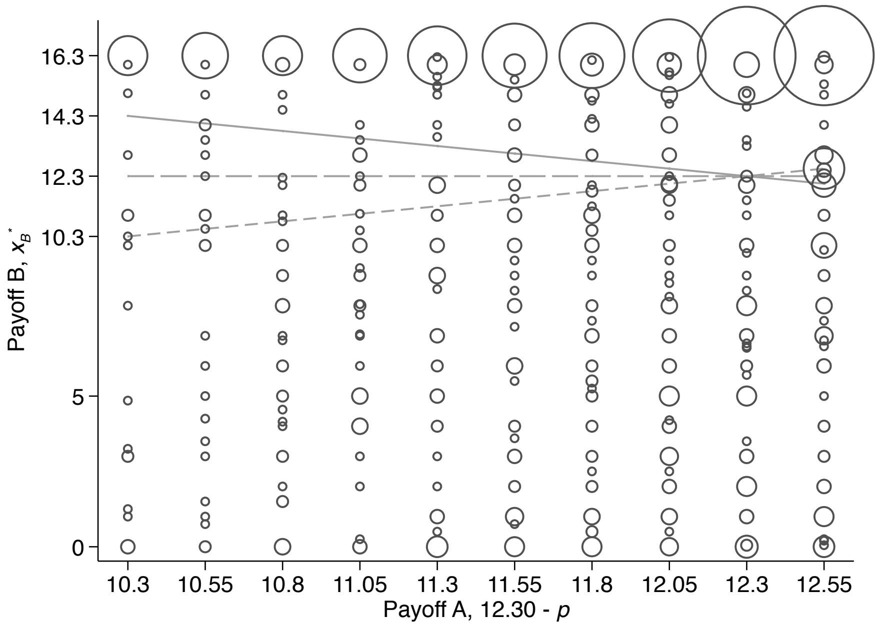

FIGURE 2. Payoff allocations $(12.30 - p, x_B^*)$ chosen by subjects in Part I.

Figure 2 clearly demonstrates that there is substantial heterogeneity in terms of the payoff allocations chosen by the subjects. The surface of each circle is proportional to the number of subjects who choose a specific allocation. For example, when price p is 25 cents, 86 subjects decide to pay and give \$16.30 to B, i.e. choose the $(12.05, 16.30)$ allocation, while when the price is 2 dollars, 25 subjects choose the $(10.30, 16.30)$ allocation. All the allocations lying above the downward-sloping solid line are efficient in that A pays less than she gives to B beyond \$12.30. More formally: $x_B > 12.30 + p$. The allocations below the upward-sloping dashed line are competitive in that A gives B less than what she receives herself, i.e., $x_B < 12.30 - p$. In other words, A decreases her own payoff in order to decrease B's payoff even further. In terms of allocation distribution, 68.2% are efficient, and the most efficient allocation of $(12.30 - p, 16.30)$ comprises 56.7% of the ones subjects are willing to pay for. Competitive allocations, where A gives less to B than she receives herself, amount to 28.8%.[13][14] About 3.4% of the allocations cannot be attributed to either category.

3.1.3. *The Power Game: Part I versus Part II behavior.* In this section we analyze subjects' behavior in the 10 subject-specific rounds of Part II.[15] In these 10 rounds, subjects are faced

[13]For the negative price of $-\$0.25$, an allocation can be both efficient and competitive at the same time, for example a $(12.55, 12.50)$ allocation. Only 0.5% or 6 out of 1,280 allocations fall into this category.
[14]Several studies have demonstrated that in different contexts people are willing to sacrifice their own payoffs in order to "burn" other people's money (Zizzo and Oswald (2001), Abbink and Sadrieh (2009), Charness, Masclet and Villeval (2014)).
[15]The remaining 12 rounds are identical for all subjects and analyzed in Section 3.3.

with choices that are determined by their decisions in Part I. For example, if in one of the rounds in Part I, a subject pays \$1 and gives \$0 to B, then it means that she prefers $(11.30, 0)$ to $(12.30, 12.30)$. In the corresponding Part II round, she faces a choice between those two allocations: $(11.30, 0)$ and $(12.30, 12.30)$. More generally, if in Part I she pays p and chooses x_B^* for B over the $(12.30, 12.30)$ allocation, then in Part II she has to choose between $(12.30 - p, x_B^*)$ and $(12.30, 12.30)$. In other words, in Part II, subjects face a choice between the allocation they chose in Part I and the alternative allocation they could have chosen but didn't.

If A retains her choice of $(12.30 - p, x_B^*)$ over $(12.30, 12.30)$ in Part II, then we say that she pays p in Part II to implement her desired allocation from Part I. Note that while the $(12.30 - p, x_B^*)$ allocation is identical to what A chose in Part I, in Part II paying p does not lead to additional power, it just leads to implementing this specific payoff distribution. Indeed, whether or not A pays p in Part II, the choice she faces is between two payoff pairs that are fixed, whereas in Part I, paying p allowed A to increase her power over B by choosing a precise payoff for B in $[0, 16.30]$. If a subject's preferences are on distributional outcomes only, she should choose the same allocations in Part II as in Part I: the subject should choose $(12.30 - p, x_B^*)$.

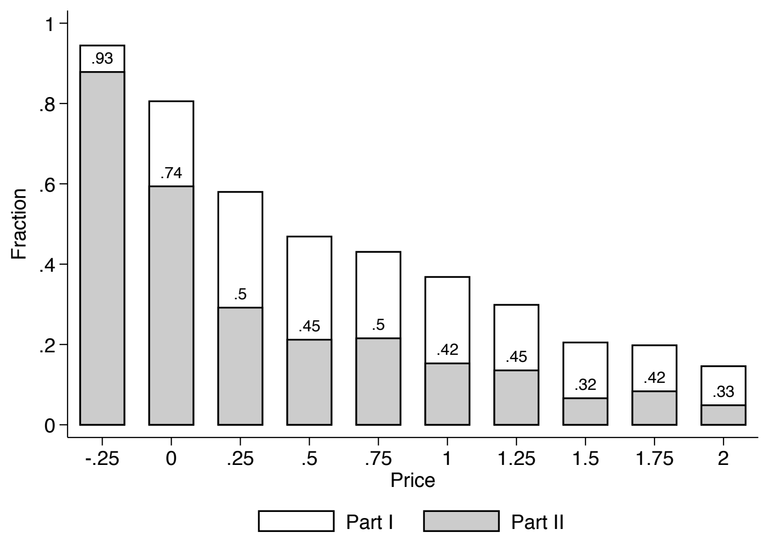

FIGURE 3. Fraction of subjects who pay in Part I and in Part II, conditional on paying the corresponding price in Part I, by price in USD.

Figure 3 adds shaded bars to Figure 1 that represent the fraction of subjects who pay in Part II conditional on paying in Part I. The conditional fraction of subjects paying in Part II for each price is written at the top of each shaded area. For example, only 32% of subjects who pay a price of \$1.50 in Part I also pay in Part II, i.e. they choose the same allocations as in Part I. There is a stark difference between subjects' willingness to pay in Parts I and II at the aggregate level for each price. For every p, conditional on having paid in Part I, in Part II far fewer subjects are willing to pay the same price to ensure the payoff distribution they chose in Part I. Interestingly, there is a difference at the price of zero, which means that some subjects prefer the $(12.30, x_B^*)$ allocation to $(12.30, 12.30)$ when x_B^* is chosen by the subjects themselves but prefer $(12.30, 12.30)$ to $(12.30, x_B^*)$ when the same value of x_B^* is fixed. Thus, even at no cost to themselves, they do not implement their Part I allocations in Part II.

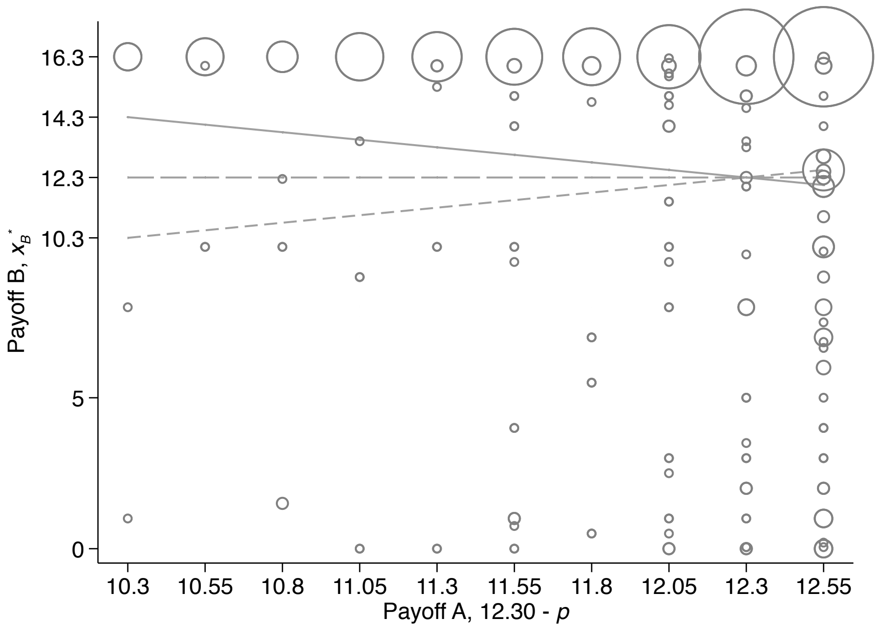

FIGURE 4. Payoff allocations $(12.30 - p, x_B^*)$ preserved by subjects in Part II.

Figure 4 shows the Part I $(12.30 - p, x_B^*)$ allocations that subjects preserved in Part II. Recall that in Part I, subjects pay p in 1,280 cases. In Part II however they do not pay in 509 or 39.8% of those cases, i.e. they revert to the $(12.30, 12.30)$ allocation. For positive prices the percentage of reversions is even higher at 55.3%. Subjects revert 23.1% and 72.4% of efficient and competitive allocations, respectively.

Our aggregate results provide strong evidence that preferences for power are non-trivial. Many subjects are willing to pay if paying increases their power over the payoffs of others as is in Part I. However, they are much less willing to pay to implement the same payoff allocations when paying does not lead to additional power as is in Part II. If subjects' decisions to obtain the right to choose payoffs of others in Part I were driven entirely by their social preferences then there should be no reversals in Part II. Thus, our results suggest that (1) preferences for power exist and are substantial and (2) that they are different than and cannot be explained by social preferences. In the next section, we continue our analysis at the individual level and explore the broad categories of preference classes among our subjects.

3.2. **Preferences for power: the individual level.**

3.2.1. *Demand functions.* In the main text we focus on those subjects who have step-shaped demand functions in part I and II of the Power Game, i.e. a single switching point. Figure 5 graphically visualizes these well-behaved demand functions. Note that Figure 5 depicts the demand functions of a subject whose willingness to pay is lower in Part II than in Part I, i.e. for whom $\bar{p}_{\mathrm{II}} < \bar{p}_{\mathrm{I}}$, but we include subjects for whom $\bar{p}_{\mathrm{II}} > \bar{p}_{\mathrm{I}}$ and $\bar{p}_{\mathrm{II}} = \bar{p}_{\mathrm{I}}$ in our analyses.

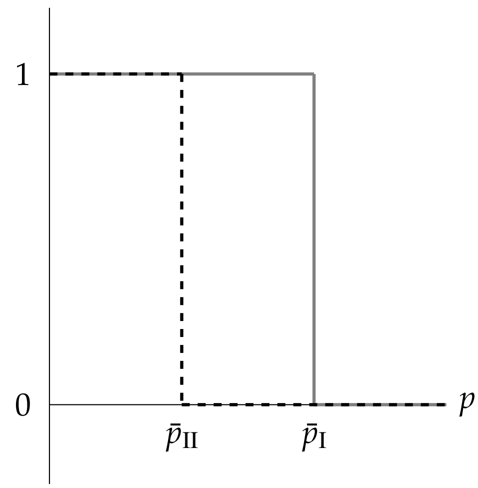

FIGURE 5. Well-behaved demand functions in Parts I and II.

We have 178 subjects (61.8% of our sample) with well-behaved demand functions in Parts I and II. This proportion is relatively large given that prices are randomly drawn in every round in Part I and that the 10 rounds that correspond to Part I are randomly presented among the 22 Part II rounds. All our results are robust to using all subjects and are not due to any selection effects. Indeed, we conduct analyses in which we allow for any number of skips and identify willingness to pay as (1) the maximum price paid; (2) the highest price paid before a decision to not pay; (3) the most consistent price, that is, the price at which a subject displays the fewest mistakes. In Appendix E, we present the results using the highest price, and all other analyses are available upon request.

3.2.2. *Difference in willingness to pay across Parts: preference classes.* Figure 6 shows the joint distribution of the subjects' willingnesses to pay in Parts I and II, $\bar{p}_{\mathrm{I}}$ and $\bar{p}_{\mathrm{II}}$. For example, for 1.69% of our well-behaved subjects, $\bar{p}_{\mathrm{I}} = 0.25$ and $\bar{p}_{\mathrm{II}} = 0.25$, while for 3.93% of our subjects $\bar{p}_{\mathrm{I}} = 2$ and $\bar{p}_{\mathrm{II}} = 0$. Subjects who are willing to pay more (less) in Part I than in Part II, i.e. for whom $\bar{p}_{\mathrm{I}} > \bar{p}_{\mathrm{II}}$ ($\bar{p}_{\mathrm{I}} < \bar{p}_{\mathrm{II}}$), appear below (above) the 45-degree line. Subjects whose willingness to pay is the same across Parts I and II lie on the 45-degree line.

We use our theoretical predictions (see Section 2) to sort subjects into different preference classes. Recall that these are only based on their willingnesses to pay in Parts I and II: different preference classes correspond to different relationships between $\bar{p}_{\mathrm{I}}$ and $\bar{p}_{\mathrm{II}}$.

Subjects with standard preferences only care about their own payoff. These subjects are never willing to decrease it to affect the payoff of others. Thus, for them $\bar{p}_{\mathrm{I}} = 0$ and $\bar{p}_{\mathrm{II}} = 0$. According to Figure 6, 47.8% of our subjects have standard preferences.

In Figure 6, subjects with power and no social preferences are located along the horizontal axis and together represent 27.5% of our sample. Indeed, in Part I they are willing to pay up to $\bar{p}_{\mathrm{I}} > 0$ to choose the payoff of B, but in Part II they never pay positive prices to implement

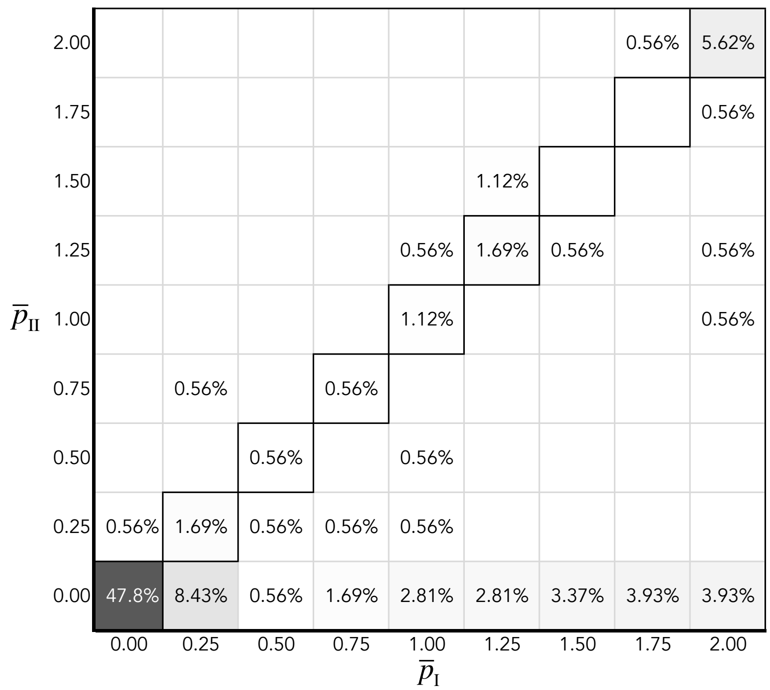

FIGURE 6. Joint distribution of the willingnesses to pay in Parts I and II, $\bar{p}_{\mathrm{I}}$ and $\bar{p}_{\mathrm{II}}$.

the allocations they chose in Part I and maximize their own payoff instead. These subjects' willingnesses to pay in Part I span the entire range of prices, from \$0.25 to \$2. In fact, 3.93% of our subjects are willing to pay up to the maximum price of \$2.

Subjects who have the same positive willingness to pay across Parts I and II of the Power Game, i.e. subjects for whom $\bar{p}_{\mathrm{I}} > 0$ and $\bar{p}_{\mathrm{I}} = \bar{p}_{\mathrm{II}}$, have social preferences and no power preferences. They derive no additional utility from power but instead care about payoff distributions, independently of how those are attained. In particular, whether they choose x_B from an interval or not has no impact on their utility. These subjects lie on the 45-degree line in Figure 6 and represent 11.2% of our sample.

Subjects who have positive but different willingnesses to pay across Parts I and II of the Power Game have preferences for power and social preferences. For 4.5% of our subjects $\bar{p}_{\mathrm{I}} > \bar{p}_{\mathrm{II}} > 0$. These subjects clearly have social preferences since they pay positive prices in Part II. However, they are unwilling to pay up to $\bar{p}_{\mathrm{I}}$ because in Part II paying does not lead to additional power. In other words, in Part I these subjects derive utility from the act of choosing a specific amount for B, as well as from the resulting distribution itself. In Part II however, they can only derive utility from the resulting distribution and so are willing to pay less. There are also subjects who pay more in Part II than in Part I: $\bar{p}_{\mathrm{II}} > \bar{p}_{\mathrm{I}} > 0$. These subjects have social preferences, dislike power and comprise 3.4% of our sample. Finally, 9 subjects, or 5.1% of our sample, never pay in Part II of the Power Game, even at the negative price of $-$\$0.25. We are unable to classify those subjects.

Figure 7 shows the distribution of preference classes among our subjects. The most common preference class is Standard: these subjects neither care about power nor about others. They

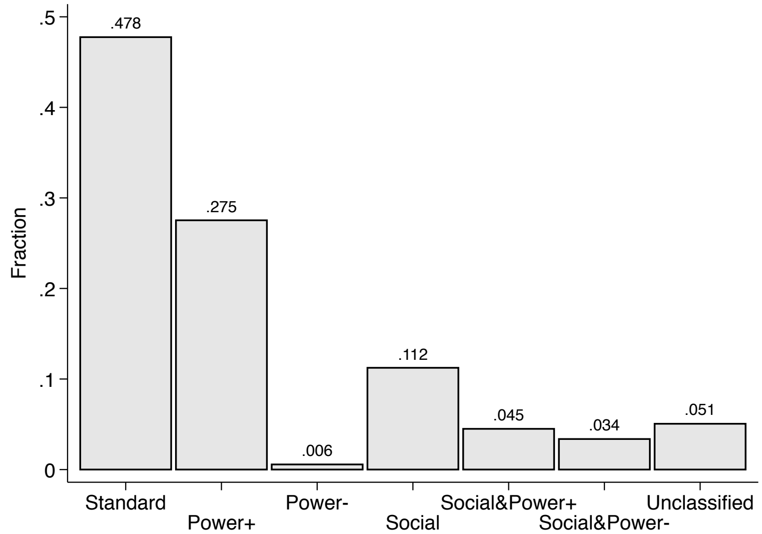

FIGURE 7. Distribution of preference classes.

represent 47.8% of our sample. The second largest class, the Power+ preference class, includes subjects who have positive preferences for power without social preferences. Such subjects represent 27.5% of our sample. Together, these two categories comprise about three quarters of the sample. Note that only one subject (0.6% of our sample) does not choose for B even when compensated to do so and therefore belongs in the Power- preference class: he or she has a negative attitude towards power and no social preferences. Subjects who have social preferences in any capacity (the Social, Social&Power+ and Social&Power- classes) represent about 19.1% of our sample. In Appendix E, we use all subjects and obtain the following distribution across preference classes: 36.1% are Standard, 25.3% are Power+ and 26.4% have social preferences, in some capacity. Therefore, the fraction of Power+ subjects is stable across samples.

Regarding how much subjects are willing to pay in order to implement their preferences, we can glean from Figure 6 that Power+ subjects are willing to pay on average \$1.08. In fact, more than half of them pay \$1.25 (about 10% of their potential payoff) or more in Part I. On average, subjects in the Social preference class are willing to pay \$1.39 to implement their preferences, while those in the Social&Power+ and Social&Power- classes are willing to pay \$1.34 and \$0.92, respectively.

3.3. **Preference classes and other behaviors.**

Our preference classification depends only on the difference in subjects' willingnesses to pay across Parts I and II of the Power Game. If our classification captures differences in preferences across subjects, then the identified preference classes should predict other subjects' behaviors. Here we provide evidence that it is indeed the case. First, we show that subjects we have classified as having social preferences, regardless of their attitude towards power, are consistent in the amounts they give to B. In contrast, subjects in the Power+ preference class exhibit much more variation in their giving behavior both within and across subjects. Second, we show that these classes also predict subjects' decisions in tasks that are unrelated to Part I of the Power Game. More specifically, in the absence of power, Power+ subjects behave much like subjects with standard preferences, that is, they maximize their own payoff, while those with social preferences do not.

3.3.1. *Choice for B's payoff.* In this section, we compare subjects identified as having social preferences and power preferences in terms of their giving behavior in rounds 1 through 10 in Part I. Subjects with social preferences belong to the following preference classes: Social, Social&Power+, Social&Power-. In other words, these subjects may be indifferent towards, like, or dislike power, but they all have preferences towards B's payoff. In contrast, Power+ subjects are indifferent towards the payoffs of others. If Power+ subjects are correctly identified, then we should see them behaving differently in terms of how they give to B compared with subjects who have social preferences.

Figure 8 shows the cumulative distribution function of the amounts given to B, averaged per subject, separately for Power+ subjects and those who have social preferences. We find that Power+ subjects on average give \$9.91 to B, compared with an average of \$16.04 for those who have social preferences.What is visually different is also different statistically.[16]

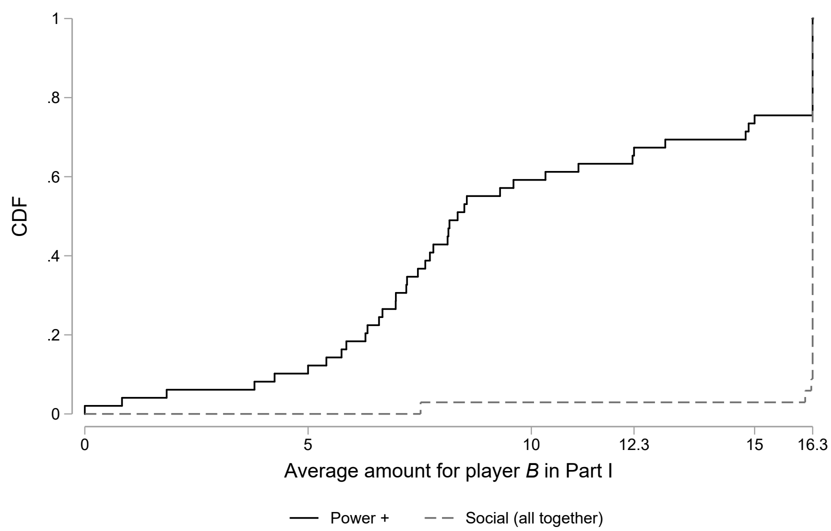

FIGURE 8. Distribution of the average amount given to player B in Part I.

Subjects with social preferences are very homogeneous: 91.2% of them always give the maximum allowable amount of \$16.30. In contrast, Power+ subjects are very heterogeneous in terms of what they give to B, and their amounts span almost the entire choice space, that is, the

[16]Kolmogorov-Smirnov and Wilcoxon-Mann-Whitney tests show that the distribution of amounts given to B by Power+ subjects is statistically different than that one of those with social preferences; both p-values are less than 0.001. The unit of observation is the average amount given to B by each subject.

$[0, 16.30]$ interval. Further, at the individual level, among Power+ subjects, there is much more variation within each subject's choice compared with subjects who have social preferences: the within-subject mean standard deviation of choices for B's payoff is higher for subjects in the former category than in the latter (2.61 versus 0.11).[17] In other words, it is not the case that Power+ subjects have strong but very different preferences towards B. Instead, they seem to have very weak preferences regarding B's payoff, as our classification implies. Similarly, this behavior cannot be reconciled with signaling models of intentions that posit individuals' desire to appear kind (see also Section 5).

We also compare subjects' giving behavior in the round in which the price is zero with behavior in round 11 of Part I. In both of these rounds, A players receive \$12.30. If subjects have social preferences, we expect them to give the same amount to B in these two rounds since their own payoff is identical in both cases. Among those subjects that we classify as having social preferences, 97.1% give the same amount in those two rounds while only 40.8% of Power+ subjects do so.[18]

The above results indicate that Power+ subjects do not try to appear consistently nice or petty and in addition attach little importance to what B players earn. This indifference towards the payoffs of others, displayed both at the aggregate and individual levels, is in sharp contrast with the homogeneous and coherent giving behavior of those identified as having social preferences. These systematic differences in giving behavior between our identified preference classes provide convincing evidence that the Power Game yields a meaningful preference classification.

3.3.2. *Behavior in a separate task.* Recall that in Part II of the Power Game, we present our subjects with 12 decision problems that are unrelated to their choices in Part I. Six of those problems (CR1 through CR6) are inspired by Charness and Rabin (2002). In Table 4 we compare our subjects to those of Charness and Rabin (2002) and Chen and Li (2009) by presenting the proportion of subjects who choose the first option in each decision problem. The way we present the decision problems in this table is such that the first option always yields a higher payoff for A, except in problem CR1 where A's payoffs are identical across the two options. The results in Table 4 indicate that our sample is largely similar to those in other institutions. If anything, our subjects seem to choose the payoff-maximizing option more often than in Charness and Rabin (2002) and Chen and Li (2009).

Importantly, as in all Part II rounds, in each of these decision problems subjects cannot increase their power by sacrificing some of their payoff, since the payoff for B is fixed in both options. That is, the amount of power subjects have is the same irrespective of which option they choose. Thus, any difference in behavior across subjects with different preference classes can only be due to their preferences beyond those for power.

[17] A two-sided test comparing these standard deviations has a p-value less than 0.001.

[18] A two-sided test of proportions confirms statistical difference, with a p-value of less than 0.001.

TABLE 4. Fraction of subjects choosing the first option in the Charness and Rabin (2002) task across three samples: Charness and Rabin (2002), Chen and Li (2009), and our sample.

Decision	First Option[a]	Second Option	CR2002	CL2009	Our Subjects
CR1	(6.60, 6.60)	(6.60, 12.30)	31%	33%	21%
CR2	(6.60, 6.60)	(6.20, 12.30)	51%	82%	57%
CR3	(10.50, 5.30)	(8.80, 12.30)	67%	76%	82%
CR4	(12.30, 3.50)	(10.50, 10.50)	27%	50%	67%
CR5	(12.30, 0.00)	(6.15, 6.15)	78%	64%	80%
CR6	(3.10, 12.30)	(0.00, 0.00)	100%	NA	99%

[a]In the experiment, what option was presented on the left or on the right side of the screen was randomly and independently determined for each subject and for each decision problem.

When additional power is not attainable, our theory (see Section 2) predicts that individuals in the Power+ preference class should behave similarly to individuals in the Standard preference class. That is, Standard and Power+ subjects should be equally likely to choose the first (payoff-maximizing) option. In contrast, subjects with social preferences should not choose the first option more often than subjects with no social preferences. Note that a subject with social preferences does not necessarily always choose the second option since her choices depend on her marginal rate of substitution between her own and B's payoff.

In Table 5 we present the fraction of subjects who choose the first option in CR1-CR6 and PT1-PT6 for each of the Standard, Power+ and Social (all together) preference classes. In order to assess our theory, we compare that fraction for the Standard and Power+ subjects as well as for the Power+ and Social (all together) subjects. Because we consider multiple outcome variables and make comparisons between two pairs of subgroups, we follow the methodology of List, Shaikh and Xu (2016)) and report multiplicity-adjusted p-values in the last two columns of Table 5. In all of what follows, we report multiplicity-adjusted p-values where applicable.

We focus first on the top section of Table 5. As is clear, across all decision problems, subjects make choices that are consistent with our preference classification and our theory. In particular, in the decision problems with a payoff/efficiency trade-off (decisions CR2-CR4, PT1 and PT2), the fraction of subjects with social preferences who choose the payoff-maximizing option is always statistically smaller than that of Power+ subjects, as the last column shows. Further, in those decisions, Power+ and Standard subjects behave similarly, as shown in the penultimate column. This is also true for CR1 in which a subject's own payoff is controlled for and CR5 where there is no efficiency gain in the second option. Aggregating behavior for each subject across decision problems CR2-CR5, and PT1-PT2, we find that the fraction of subjects who always choose the payoff-maximizing option among those with standard preferences is 52.9%. For Power+ subjects this fraction is 55.1%, statistically indistinguishable from those in the

TABLE 5. Fraction of subjects choosing the first option in the independent decision problems in Part II by preference class.

	Option		Preference Class			p-value	
Decision	First	Second	Standard	Power +	Social (All)	P+ vs. Std	P+ vs. Soc
CR1	$(6.60, 6.60)$	$(6.60, 12.30)$	20%	31%	3%	0.793	0.005
CR2	$(6.60, 6.60)$	$(6.20, 12.30)$	68%	67%	12%	0.910	0.000
CR3	$(10.50, 5.30)$	$(8.80, 12.30)$	91%	94%	50%	0.940	0.000
CR4	$(12.30, 3.50)$	$(10.50, 10.50)$	76%	84%	26%	0.880	0.000
CR5	$(12.30, 0.00)$	$(6.15, 6.15)$	82%	88%	71%	0.974	0.833
PT1	$(10.10, 5.20)$	$(9.10, 9.10)$	78%	84%	32%	0.963	0.000
PT2	$(12.30, 5.10)$	$(10.10, 12.30)$	86%	92%	29%	0.870	0.000
Sanity checks							
CR6	$(3.10, 12.30)$	$(0.00, 0.00)$	100%	98%	100%	0.945	1.000
PT3	$(12.55, 12.80)$	$(12.30, 12.30)$	96%	96%	97%	0.998	0.987
PT4	$(12.30, 9.60)$	$(9.60, 12.30)$	99%	100%	91%	0.961	0.832
PT5	$(12.30, 7.80)$	$(7.80, 5.40)$	100%	100%	100%	1.000	1.000
PT6	$(6.15, 6.15)$	$(0.00, 0.00)$	100%	100%	100%	1.000	1.000

Standard preference class, and it is significantly higher than 5.9%, the fraction for those who have social preferences.[19]

We grouped CR6, PT3-PT6 together as these questions provide a test for whether subjects cared about their payoffs and understood our instructions. For each of these "sanity checks," all or almost all subjects choose the first option, irrespective of their preference class. Importantly, almost all subjects choose the first option in PT4, which allows us to show that subjects understand that they are to act as type A players (if this were not the case, more subjects would have chosen the second option). Decisions in CR6, PT3, PT5, and PT6 provide evidence that subjects care about their own payoffs and also do not choose payoff-pairs for which both players receive lower payoffs. More generally, the very high level of first-option choices across all preference classes in the "sanity check" problems confirms that subjects' behavior in the Power Game cannot be explained by confusion regarding roles or payments.

3.4. **Stability of preference classes.** Here we investigate the stability of preference classes. In order to do so, we use data from Part I* (the final task where subjects play Part I of the Power Game for the second time) to re-classify our subjects. We employ the same preference class definitions as in Section 3.2 but use subjects' paying behavior in Parts I* and II, instead.

Figure 9 presents the distribution of preference classes based on subjects' paying behavior in Parts I* and II. As is clear from Figures 7 and 9, the overall fractions of subjects in the different preferences classes are only marginally affected, if at all, by whether we use Part I or Part I* for the classification. We cannot reject the null hypothesis that the fraction of subjects in each preference class using Parts I* and II is the same as the one identified using Parts I and II (all the p-values are above 0.10).

[19]The p-value for the first two is 0.806, while the p-values for each pairwise comparison between social and the others are smaller than 0.001.

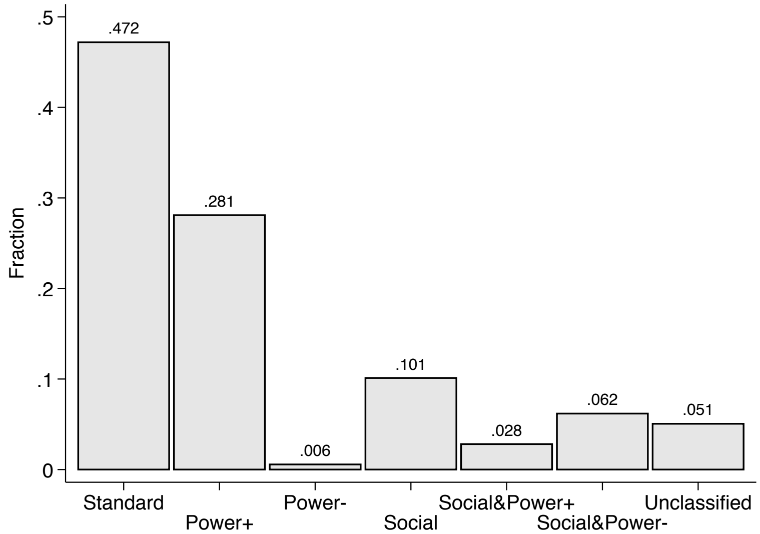

FIGURE 9. Distribution of preference classes as obtained from Parts I* and II.

However, aggregate stability could potentially hide individual heterogeneity. We show that this is not the case. First, the overwhelming majority of our subjects, or 84.3%, retain their preference class between Parts I and I* of the Power Game. Then, to address the issue of individual-level stability, we examine, for each preference class defined using subjects' paying behavior in Parts I and II, the fraction of subjects who retain that preference class defined using Parts I* and II of the Power Game. Figure 10 displays the results.

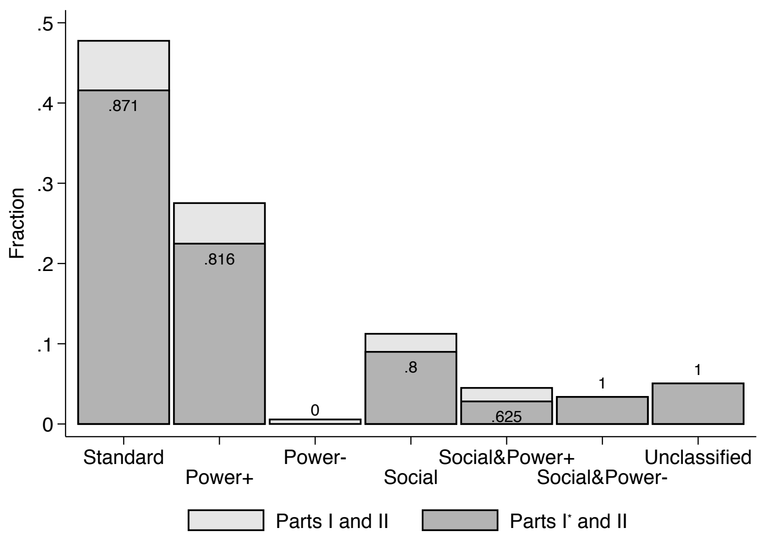

FIGURE 10. Fraction of subjects in each preference class, as obtained from Parts I and II, who retain their preference class, as obtained from Parts I* and II.

As Figure 10 shows, stable subjects represent the majority in each preference class.[20] Among subjects identified as having Standard preferences using Parts I and II, 87.1% retain their class.

[20]The exception to this is the lone Power- subject who disappears when using data from Parts I* and II.

Among Power+ subjects and subjects with social preferences (all together), these fractions are 81.6% and 79.4%, respectively. We cannot reject the null hypothesis that the fractions of subjects with stable preferences are the same for all preference classes (all the p-values are above 0.10).

Finally, we show that the strength of subjects' preferences does not significantly change across Parts I and I*. To demonstrate that we compare subjects' paying and giving behavior in Parts I and I* for those subjects with stable preferences for each preference class. Our findings are presented in Table 6.

TABLE 6. Subjects' behavior in Parts I and I*, for those subjects who retain their preference class in Part I*.

Panel A: Subjects' paying behavior

Preference Class	Subjects	Max Paid Part I	Max Paid Part I*	Paid the Same in Parts I and I*
Standard	74	0.00	0.00	1.00
Power +	40	1.24	1.17	0.78
Social (All)	27	1.29	1.26	0.81
Social	16	1.34	1.34	1.00
Social & Power +	5	1.55	1.40	0.40
Social & Power −	6	0.92	0.92	0.67

Panel B: Subjects' giving behavior

Preference Class	Subjects	Payoff for B Part I	Payoff for B Part I*	Gave the Same in Parts I and I*
Standard	74	14.22	14.41	0.80
Power +	40	9.41	10.04	0.32
Social (All)	27	16.00	15.98	0.97
Social	16	16.30	16.30	1.00
Social & Power +	5	14.68	14.60	0.84
Social & Power −	5	16.30	16.30	1.00

Panel A of Table 6 presents subjects' willingness to pay to implement their preferences in Parts I and I*, as well as the fraction of subjects who did not change that willingness across the two Parts. For all preference classes, at the aggregate level, subjects' willingness to pay is very stable between Parts I and I*.Not only are average willingnesses to pay across Parts I and I* very similar, but even at the individual level, a significant majority of subjects don't change how much they are willing to pay to implement their preferences, as the last column in Panel A shows. In addition, there is no systematic direction subjects change their willingness to pay.[21]

Panel B of Table 6 displays subjects' giving behavior in Parts I and I*. Taking each preference class in turn, for each subject we calculate the average amount given to player B in all rounds when she chose for B, including Round 11 of Part I (I*), and then average this over all subjects in that preference class. At the aggregate level, how subjects in different preference classes give to B varies very little across Parts I and I*.

[21]For Standard, Power+ and Social (all together) preference classes, the p-values are greater than 0.10.

In the last column of Panel B we present the fraction of subjects who, for all rounds when they choose for B, have identical giving behavior in Parts I and I*. Our theory predicts that subjects with social preferences should give the same amounts to B when their own payoff is controlled for. This is precisely what we observe in the data at the individual level: 97% of subjects with social preferences give the same amounts across all prices at which they pay. In other words, they behave identically in Parts I and I*. While our theory does not impose any restrictions on the process by which subjects in the Standard or Power+ classes choose B's payoffs, it is reasonable to think that how much they give to B would be less stable across Parts I and I*, compared with subjects with social preferences. Only 32% of Power+ subjects have identical giving behavior in Parts I and I*, a significantly lower fraction than for those with social preferences.[22] So while the average Power+ subjects give to B remains stable, the specific amounts they give vary across Parts I and I* even when their own payoff is controlled for.[23] This provides evidence that their intent is to vary the payoffs of others, but not necessarily in any specific direction. Similarly, 80% of subjects in the Standard class have identical giving behavior in both Parts I and I*, and while this fraction is higher than for those in the Power+ class, it is still significantly below that of the Social (all together) preference class.[24]

4. Discussion

In this section we discuss potential mechanisms underlying individuals' preference for power. We conducted 4 additional treatments in order to improve our understanding of our subjects' behavior in the main experiment. These four additional treatments consist of 3 "modified" Power Game experiments that parallel our main design and one set of vignette studies. These additional treatments were conducted in February 2019.

4.1. The modified Power Games.

In all the modified Power Games, our theoretical predictions are valid for Standard subjects, and whether they hold for Power+ subjects depends on whether they value a specific aspect of power or not. Table 7 shows the fraction of Power+ subjects as well as their willingness to pay in Part I of the Power Game in all four treatments (the three modified Power Game treatments as well as the main treatment).

In the first additional treatment, the "12.30 treatment" the only difference with the main experiment is that when A players pay in Part I, they can give B an amount between \$0 and \$12.30. The 12.30 treatment is designed to address two questions: (1) Is the ability to be kind (i.e. give more than \$12.30) or petty (i.e. give less than \$12.30) important for the individuals with preferences for power in our main experiment? and (2) Does any reduction in choice space negatively impact the value of power?

We collected data from a total of 82 subjects in the 12.30 treatment. Out of these, 48 (or 58.64%) are well-behaved. The fraction of Power+ subjects in the 12.30 treatment is 20.83%,

[22] The p-value for a two-sided test of proportions is less than 0.001.

[23] Thus, we find no evidence that "power corrupts" (Bendahan et al. (2015)). However, it is possible that in our setting subjects did not experience power for a long enough period of time for such an effect to take place.

[24] The p-value for a two-sided test of proportions is 0.0106.

which is statistically no different than in the main treatment. However, their willingness to pay is significantly lower than in the main treatment: here the Power+ subjects are willing to pay \$0.60 on average, compared to \$1.08 in the main treatment. Our results show that subjects still enjoy their ability to determine outcomes of others even when their choice space is constrained. However, these subjects recognize this constraint and value power in the 12.30 treatment less than in the main treatment. One can imagine that a further shrinking of the choice space would lead to a further reduction in willingness to pay and a smaller number of Power+ subjects.

TABLE 7. Power+ subjects across different treatments.

	Main	12.30	Charity	Computer
Total number of subjects	288	82	82	100
Percent of well-behaved	61.81	58.54	60.98	47.00
Percent of Power+	27.53	20.83	4.00***	2.13***
WTP of Power+ in Part I, $\bar{p}_{\mathrm{I}}$	1.08	0.60**	0.25	0.75

In the second setting, the "Charity treatment," subjects can again give anywhere between \$0 and \$16.30 (as in the main experiment), but they choose how much to give to a charity (Four Diamonds/THON) not to another player. We recruited an additional 82 subjects for this treatment, 50 (or 60.98%) of which were well-behaved. As is clear from Table 7, we observe a drastic decrease in the fraction of Power+ subjects. This provides evidence that subjects who pay in the main treatment (or the 12.30 treatment for that matter) do not do so simply because of the "lure of choice" (Bown, Read and Summers (2003)) since in the Charity treatment their choice space is the same as in the main treatment. This result also suggests that the distance between an individual making decisions and the "other," as well as the impact that the decision-maker can have on the "other" may matter in the perception and valuation of power.

Finally, in the third, "Computer treatment," when player A chooses to pay, a computer randomly selects B's payoff, uniformly between \$0 and \$16.30. For this treatment, we recruited a total of 100 additional subjects, 47 of which were well-behaved. In this treatment, Power+ subjects virtually disappear. This suggests that Power+ subjects do not simply desire random payments for B, as one may think might be the case when looking at Figure 8.[25] In contract, these results highlight the importance of being able to determine payoffs of others directly as opposed to simply influencing them. These results also relate to work by Ferreira, Hanaki and Tarroux (2017) and Neri and Rommeswinkel (2017), who focus on situations where subjects can only influence payoffs of others in a probabilistic way and find that subjects there are not willing to pay for the ability to influence. These findings are also in line with Chassang and Zehnder (2019), who, in a different context, show that subjects are willing to make certain decisions that negatively impact others when actions have direct and clear repercussions, but that they are less willing to take these actions when there is more noise in the environment.

[25]For example, Agranov and Ortoleva (2017) show that individuals may have preferences for randomization over the choice of lotteries in certain settings.

4.2. **Vignette Studies.** In addition to the laboratory experiment, we also conducted a vignette study to illustrate individuals' perception of power by third parties in different real-life situations. An additional 102 subjects participated in the vignette study. In these vignettes, subjects were exposed to 5 hypothetical everyday workplace and university-life scenarios.

In three of the scenarios we vary the type of the choice set (singleton versus an interval) for the decision maker. In the other two scenarios, we vary whether information on who determines outcomes is fully revealed. The description of each scenario is identical except for the two conditions. The exact text of our scenarios is available in Appendix F. Each subject goes through all five scenarios but sees only one condition for a given scenario. After reading each scenario subjects are asked to identify how much power an individual in a scenario has over other individuals on a five-level scale, ranging from "1 = No power at all" to "5 = Very high amount of power."

- In scenario 1, a CEO collects proposals from his employees and decides on the reward mechanism. In Condition 1, he offers a fixed reward to everyone, while in Condition 2 he carefully reviews the proposals and offers a variable reward. In both conditions the subjects are asked about the amount of power the CEO has over the rewards for his employees.
- In scenario 2, a professor is deciding how to evaluate his students. In Condition 1, he opts for pass/fail and in Condition 2, for numerical grades from 0 to 100. The subjects are asked about how much power the professor has over the grades of his students.
- In scenario 3, a manager decides on bonuses for her employees. In Condition 1, she recommends the same bonus to all her workers, in Condition 2, she recommends different bonuses depending on performance. The subjects indicate how much power the manager has over the bonuses of her employees.
- In scenario 4, a professor asks his TA to grade students' exams. In Condition 1, he informs the students that their exam will be graded by his TA. In Condition 2, he doesn't inform them. The subjects are asked how much power the TA has over the students' grades.
- In scenario 5, a manager is about to promote one of his workers and makes a recommendation to his boss. His boss reviews the candidates and has final say. In Condition 1, his boss informs the workers of her role in the promotion decision. In Condition 2, she doesn't inform them of her role. The subjects are asked how much power the manager has over the promotion decision.

Our results are presented in Table 8 and provide suggestive evidence that our experimental results obtained in a controlled lab environment are applicable to realistic everyday interactions between individuals. There is a strong statistical difference in how subjects perceive the amount of power across conditions for Scenarios 1 and 3, the two that most resemble our main experiment. Indeed, in both cases, when a CEO/manager is able to choose from an interval when deciding on a payment, she has a much larger amount of power than when the payment amount is fixed. This provides support to the idea that in the Power Game, A can increase her power when she chooses B's payoff from an interval.

TABLE 8. How much power does a decision-maker have?

	Condition 1	Condition 2	p-value
Panel A: Fixed vs. variable reward			
S1: CEO assigns rewards for proposals	2.66	4.04	0.0003
S2: Professor grades a course	3.51	3.48	0.9227
S3: Manager recommends end-year bonuses	3.06	4.06	0.0003
Number of subjects	53	49	-
Panel B: Known vs. hidden intervention			
S4: TA grades final exam instead of professor	3.49	3.22	0.5523
S5: Manager's recommendation overseen by boss	3.11	3.18	0.9193
Number of subjects	53	49	-

However, subjects don't simply equate choice set size with power. Indeed, our above conclusions don't carry over to the grade domain: in Scenario 2, we find no statistical difference between our two conditions. In other words, power is increased when an individual can choose from an interval when it comes to payments, but not when it comes to grades. Thus, whether power varies with choice set size likely depends on the domain.

In scenarios 4 and 5 we focus on whether information on who determines outcomes impacts the perception of power. The data from scenario 5 show that a manager's power doesn't seem to be affected by whether or not employees know that the manager's boss has final say. This provides support for the idea that in our main experiment, the power is not diminished by the fact that subjects don't know what Part or Round of our experiment determined their payment.[26] That is, we do not reduce A's perception of power by the fact that her choices may not end up being implemented.

The data from scenario 4 show that this is also true in the grade domain: the TA determining grades has as much power over the grading whether or not students in the class know that the TA is doing the grading. Thus, both the payment and grade domains show similar conclusions in terms of how information on possible external interventions impacts the perception of power.

5. ALTERNATIVE EXPLANATIONS

We discuss whether the behavior that we identify as evidence of preferences for power may in fact be due to other factors. Since Power+ subjects pay in Part I and not in Part II, we must consider other possible motivations that would lead to such behavior.

[26] A players here resemble the manager, the randomness of which Part gets chosen for payment resembles the boss intervening, and the employees are B players.

5.1. **Intentions-based social preferences.** Here, we argue that neither intentions-based (IB) social preferences[27] nor the interdependent preference models[28] can account for the behavior of Power+ subjects. In such models, an agent may act in a certain way in order to achieve her objectives: to signal her type, to avoid feelings of guilt, to reciprocate, or to react to the type of subject she believes she is matched with. One may be tempted to use such theories to explain away our paper's conclusions by establishing a different channel, one not related to power, to account for a subject's decision to pay in Part I and not in Part II. Below we discuss several such theories, show that they are inconsistent with our data, and more generally argue that IB and interdependent preference models cannot explain Power+ subjects' behavior.

5.1.1. *Design features used to minimize the potential impact of IB preferences.* Given the broad set of IB theories as well as the large body of evidence that these theories do characterize some individuals' behavior in other contexts,[29] we designed our experiment to minimize their potential impact. In order to do so, we turned to the literature that has investigated when such preferences are more or less salient (see, for example, McCabe, Rassenti and Smith (1998), Blount (1995), Charness and Levine (2007), Sebald (2010), Rand, Fudenberg and Dreber (2015), Toussaert (2017)). The findings are that the reciprocal response from an agent is typically much weaker when outcomes cannot be attributed to particular actions or intentions or in the presence of asymmetric information as to what alternative options other players face. Using these results, we made sure that subjects in our experiment could not link their final payoffs to intentions or actions of others and that this feature was common knowledge. We also maintained information asymmetry in terms of choices each subject faces.

Three aspects of our design allow us to separate final payoffs and actions. First, we incorporated an initial task, the Lottery task, which severed the link between payments and actions (see Section 1.3). Indeed, any final payoff a subject obtained in our experiment may have been an outcome of the Lottery task and not the result of an individual's decisions in the Power Game. In this sense, our Lottery task resembles the "random devices" that have been used in extensive form games (Charness and Dufwenberg (2006), Rand, Fudenberg and Dreber (2015), Toussaert (2017)).

In Table 9 we show the fraction of subjects who preferred the lottery over the fixed option for each of the fixed options. Given the proportion of subjects choosing the lottery in the various rounds of the Lottery task, subjects could reasonably assume that a payment of \$12.30 (the median earnings in our experiment), \$16.30 (the second most common earnings in our experiment), or anything in between, might have come from this task, as opposed to from an A

[27] See, for example, Geanakoplos, Pearce and Stacchetti (1989), Rabin (1993), Dufwenberg and Kirchsteiger (2000, 2004), Falk and Fischbacher (2006), Segal and Sobel (2007), Battigalli and Dufwenberg (2007, 2009).

[28] See, for example, Levine (1998), Cox, Friedman and Gjerstad (2007), Cox, Friedman and Sadiraj (2008), Gul and Pesendorfer (2016). See also Sobel (2005) for a related discussion on these literatures.

[29] See, for example, Dufwenberg and Kirchsteiger (2004) Charness (2004), Charness and Dufwenberg (2006), Falk, Fehr and Fischbacher (2003, 2008), Vanberg (2008), Ederer and Stremitzer (2017).

player. Consequently, no player could infer the actions of another player, therefore could not assign intentions, and all players knew this.

TABLE 9. Fraction of subjects choosing the [$0, $16.30] lottery in the Lottery task.

Fixed option	$0	$3.10	$6.60	$12.30	$16.30
Fraction choosing the lottery	100%	97.59%	73.49%	0.60%	0.00%

Second, we started the experiment by informing subjects that at the conclusion of the experiment they would only know their own payoff and would be given no other information. Thus, subjects were aware that they would not be given any information regarding their "true" types, nor would they be told which task or round was chosen for payment. Together with the Lottery task, this lack of information represented a significant obstacle to assigning intentions to other players' actions based on observable final payoffs.

Finally, in both Parts of the Power Game we maintained asymmetry in information as to what alternatives subjects faced. In Part I this asymmetry came from the fact that the set of prices a subject faced was not public information and subjects could reasonably assume that it varied from subject to subject. In Part II, subjects were given no information about payoff pairs the others faced, so they could also reasonably assume that they may have differed across subjects.

5.1.2. *Our data are incompatible with intention-based models.* Several studies emphasize individuals' desire to appear "nice" and to signal their good type to others (Geanakoplos, Pearce and Stacchetti (1989), Charness and Levine (2007), Toussaert (2017)). In our setting, some subjects may feel that they are in a better position to signal their good intentions in Part I than in Part II. In Part I, it is common knowledge that A can increase B's payoff if she pays p, even though subjects do not know the range of p for all players. In Part II, however, A realizes that others do not necessarily know her options and therefore she cannot signal her good type to B. Such subjects would pay in Part I, give $16.30, but wouldn't pay in Part II. Contrary to this argument, Power+ subjects give unsystematically to B (see Figure 8 and Table 6), i.e. they do not try to appear "nice" in Part I. One might also imagine that subjects want to signal their good intentions/type to the experimenter who observes their trade-offs in Parts I and II. However, such signaling would imply consistent paying and giving behavior in both Parts. If anything, this argument might explain behavior of Social subjects, who pay in both Parts and consistently give high amounts to B. In addition, in the 12.30 treatment subjects cannot be generous, and yet we find a similar proportion of Power+ individuals.

Likewise, the hypothesis that subjects' guilt (Charness and Dufwenberg (2006), Ellingsen et al. (2010)) generated by their unsystematic and often unkind giving behavior in Part I motivates their choices in Part II is inconsistent with our data. First, in Part II Power+ subjects revert *all* their choices, including the ones where they had given high amounts to B in Part I. Then, in those Part II rounds that are independent of Part I, they do not try to make up to B and instead favor payoff-maximizing options, even when they can greatly increase B's payoff at a very low cost to themselves (see CR1 and CR2 in Table 5). Finally, Power+ subjects return to their unsystematic giving behavior in Part I*.

Another set of explanations relies on the idea that strategies are evaluated relative to other strategies available to the decision maker (Falk, Fehr and Fischbacher (2003), List (2007)). For example a subject may feel particularly good about herself when in Part I she pays p and gives \$16.30 because she knows she could have given \$0 instead. In Part II, she may not get as high an additional utility because she chooses between $(12.30, 12.30)$ and $(12.30 - p, 16.30)$ and the lowest available payoff for B is \$12.30 not \$0. However, such Power+ subjects would then give \$16.30 to B in Part I, a choice that is not widespread in our data.As a second example, a subject may give any strictly positive amount $x_B > 0$, and, because it is better than the lowest alternative of \$0, still feel good about herself, even though she's not giving the maximum. In Part II, these numbers are compared with \$12.30 and may no longer appear kind. In both cases, such subjects would revert choices that appear unkind in Part II, i.e. those where $x_B <$ \$12.30, and retain those that appear kind, i.e. those where $x_B > 12.30$. Power+ subjects, however, revert all their decisions. In any case, for a given price, subjects who evaluate strategies relative to other strategies should then give the same amounts in Parts I and I* (since Parts I and I* are identical), which only a minority of Power+ subjects do.

More generally, explanations that take into account the actual choice of x_B as part of their logic (either on its own or relative to another parameter) cannot explain the behavior of Power+ subjects. Indeed, if Power+ subjects incorporated B's payoff in their own utility, this would, presumably, have been done in a systematic and predictable way. Our data show the opposite: in Part I, Power+ subjects choose to give in a very unsystematic way, sometimes giving very high amounts, sometimes very low ones, with no particular pattern for a given subject (see Table 5, Panel B of Table 6 and Figure 8). Thus, the choice of what to give to B is unlikely to be the result of a maximization process, which is also made clear by the fact that those subjects don't give the same amounts across Parts I and I* for a given price.

5.2. **Preferences for randomization.** One might argue that Power+ subjects pay in Part I because they have preferences for randomizing payoffs of B players. Our results are incompatible with this hypothesis. First, in the computer treatment we find only a negligible fraction of Power+ subjects. Second, in our vignette studies, we find that individuals consistently perceive being able to choose someone's compensation from an interval as having more power compared to implementing just one option.

5.3. **Focusing on well-behaved subjects, mistakes in paying behavior and confusion.** In the main text we focus on well-behaved subjects, i.e. those who make no skips in Parts I and II of the Power Game. We also re-do our analyses using the entire sample of 288 subjects. For that we need to adjust our definition of willingness to pay. If a subject makes no skips, her willingness to pay is calculated as before. If a subject skips some prices, her willingness to pay is defined as the maximum price p at which she pays before making her first skip, or $-$\$0.25 if she does not pay at a price of $-\$0.25$.[30] Using the entire sample, we re-do all our Tables and Figures, and present the results in Appendix E. All our conclusions are unchanged. In addition, as can be seen in the last row of Table 10, we can reject the argument that Power+ subjects are simply those who make more mistakes. In fact, they make fewer skips than the subjects with social preferences (all together).

TABLE 10. Confusion and skips in Parts I and II of the Power Game by preference class.

Preference Class	All	Standard	Power+	Social (all)	Unclassified
Panel A. Confusion in the well-behaved sample					
Number of subjects	178	85	49	34	9
Fraction confused	0.12	0.13	0.08	0.06	0.44
Panel B. Confusion and skips in the whole sample					
Number of subjects	288	104	73	76	32
Fraction confused	0.18	0.14	0.15	0.14	0.44
Number of skips in Parts I and II	1.20	0.66	0.82	1.37	3.09

We also show that Power+ subjects are not more confused about the game than others. In the final questionnaire, we asked subjects whether they found anything in the experiment confusing (see Appendix C). Table 10 presents the fraction of subjects whose answer was positive, regardless of what they indicated was confusing. In both the sample of well-behaved subjects (Panel A) as well as the entire sample (Panel B), the fraction of subjects who indicated that they found some aspect of the instructions confusing is stable and not statistically different across preference classes. The only exception is the "unclassified" category, where subjects were significantly more likely to be confused. Moreover, in the well-behaved sample the level of confusion is far lower than in the entire sample.

5.4. **Mistakes in giving behavior.** One might also argue that Power+ subjects are mis-identified and instead have social preferences. Those preferences lead them to pay in Part I but then they make mistakes in terms of what to give to B. They might realize their mistakes and revert all their choices in Part II, appearing to us as though they belong in the Power+ preference class. Two elements refute this hypothesis.

First, recall that over 81% of the subjects in the Power+ class retain their class in Part I* and pay again in order to choose for B (see Figure 10). So if they realized payoff mistakes in Part II, they should correct them in Part I*. However, their giving behavior in Part I* is remarkably close to their behavior in Part I: they continue giving unsystematically. To illustrate this, we analyze subjects' giving behavior in Part I* and re-do Figure 8 for subjects who retain their preference class in Part I*. We compute the cumulative distribution function of the amounts given to B in Part I*, averaged per subject, separately for Power+ subjects and those with social preferences, and present the results in Figure 11.

[30] We also can define a subject's willingness to pay as the maximum price she ever pays. Our conclusions are unchanged if we do so. These results are available upon request.

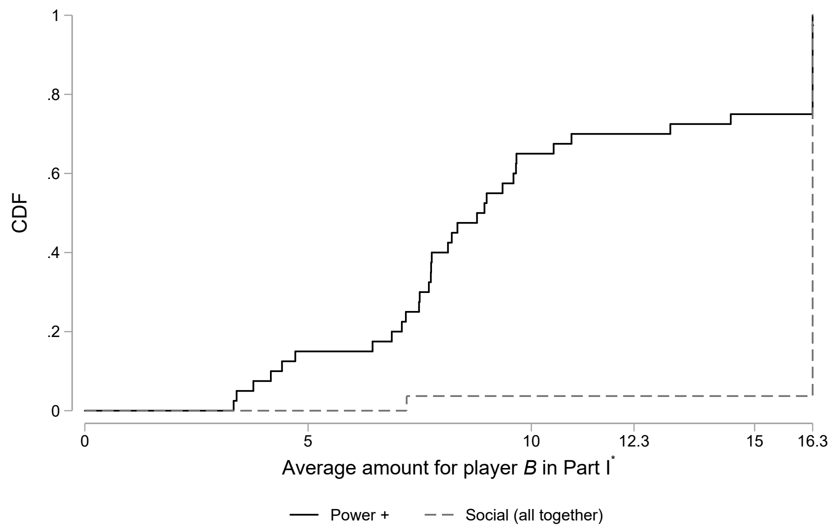

FIGURE 11. Distribution of the average amount given to player B in Part I*.

Figure 11 shows that there are clear differences between subjects in terms of what they give B in Part I* and that these differences are consistent with their preference class.At the individual level, the within-subject mean standard deviation of choices for B's payoff is higher for Power+ subjects than for subjects with social preferences (3.02 versus 0.15, $p < 0.001$). Thus, subjects in the Power+ preference class continue to exhibit substantial variation in their giving behavior in Part I*.

Second, we can use behavior in the independent choice problems in Part II. Since payoffs in those independent choice problems are separate from any choice they made in Part I, decisions Power+ subjects make in these questions would resemble those of subjects in the Social classes if in fact those were their "true" preferences. Instead, Power+ subjects resemble subjects with standard preferences and are very different from those with social preferences.

5.5. **The Lottery task influences behavior and experimenter demand makes subjects pay in Part I.** Prior to the experiment described and analyzed in this paper, we had run an experiment with a different design. We had 16 sessions and 292 subjects in total. In the earlier version, we ran only the Power Game, Parts I and II, that is, only tasks 2 and 3 of the current setup. In particular, subjects did not play the Lottery task or repeat Part I of the Power Game. In terms of interface, in Part I, subjects' screens were also different. In a first stage, they faced a single question: "Do you wish to pay $\$X$ to choose for B?" and had to answer "yes" or "no." In the second stage, if they answered "yes," the screen they faced consisted of text and a box to input the amount they wished to give to B. If they answered "no," subjects had to choose their own payoff in the $[0, 12.30]$ interval, as opposed to receiving $12.30 and having to enter a sequence of 1 to 5 characters. Finally, there was no negative price and subjects were told they would find out their "true" types at the end of the experiment, which they did.

The impetus for running a new design was threefold: the Lottery task allows us to address the issue of intentions-based social preferences and interdependent preference models; having a task after Part II resolves potential problems related to Part II being the last task; and finally, running Part I twice allows us to show that preference classes are stable.

There are no substantive differences in our results across both designs: Power+ subjects represent, in both cases, a substantial fraction of the population; These subjects give notably different amounts compared to subjects with social preferences; Preference classes are predictive of choices in independent decision problems (see Appendix G for details). Most importantly, the robustness of our findings across two different experimental implementations allows us to address elements of the current design that potentially could have made some subjects appear as though they have power preferences when in fact they do not.

5.5.1. *The Lottery task influences behavior.* One may worry that having the Lottery task makes some people think they need to choose for B and give randomly. In Part II of the Power Game, this is no longer an issue and so they revert to the payoff-maximizing option. If this is the case, then Power+ subjects are simply those who were influenced by the Lottery task. In our previous design, there was no Lottery task, yet Power+ subjects gave just as irregularly.

5.5.2. *Experimenter demand and boredom.* One may argue that when a subject doesn't choose for B, their task in the second stage is of no consequence for payoffs, and this may push subjects to pay. If they do not pay, they may find entering 1 to 5 characters to be less exciting. They may also feel pressured to do something "that matters." In the previous design, subjects who chose not to pay had to make decisions that directly determined their own payoff. Thus, since in the earlier design both the choice to pay or not lead to subjects determining payoffs (either their own or someone else's), experimenter demand as well as boredom were less present. Yet Power+ subjects were still a substantial fraction of the population. Further, De Quidt, Haushofer and Roth (2018) show that experimenter demand is typically weak.

5.6. **Uncertainty regarding type assignment.** Recall that in our experimental implementation, subjects are never informed about their true type but instead are asked to make decisions as if they were type A players. Subjects' uncertainty about their type might create two potential problems. The first is that it might generate confusion and lead subjects to incorrectly believe that their decisions might matter for their payoff even if their true type turns out to be B. Such a subject would incorrectly believe that if she pays, she receives $\$12.30 - p \text{ or } x_B^*$, depending on her realized type, and receives \$12.30 for sure if she doesn't pay. The second potential problem is that not knowing what type one is might make a subject more likely to empathize with a type B player, and possibly exacerbate her social preferences (Brandts and Charness (2011)). However, if either of these two factors affects a subject's decisions in Part I, it should also affect her decisions in Part II in the same way. In other words, these subjects' paying behavior should be identical across Parts. Thus, type uncertainty has no impact on our identification of power preferences, but might lead us to over-identify social preferences relative to standard preferences compared with a design with no type uncertainty.

5.7. **Time trends and price ordering.** One may argue that our results are simply due to time trends, or to the fact that subjects only face positive prices up until the very end of Part I of the Power Game. Time trends may be an issue if, for example, individuals' power or social preferences can be satiated over time. Several elements rule out this possibility. First, the instructions were very clear that only a single round in the experiment would be chosen for payment. Thus, actual power or generosity can only happen if a subject consistently implements her preferences in every round. Then, in our main text and analyses we focus on individuals who have well-behaved demand functions. This de facto controls for time. Further, in Appendix E we show that our results are robust to allowing any number of skips. Further, we have shown that subjects' preferences are stable across Parts I and I*. Finally, both probit and logit regressions, whether clustering at the session or individual levels, show that while the decision to pay is highly negatively correlated with price, it is not impacted by time. Thus, our findings are not due to time trends.

Similarly, we refute the argument that subjects pay in Part I because they face positive prices and that if they had faced a negative price early on they wouldn't pay positively. When subjects play Part I* they have already faced a negative price. Yet, they behave in a way that is consistent with their Part I behavior. Thus, paying behavior in Part I is not due to only facing positive prices until the very end.

5.8. **Ordering of Parts I and II.** One may wonder whether our results are driven by the fact that in the Power Game subjects always face Part I before Part II. It is possible that Part I fundamentally changed the subjects and therefore affected their Part II behavior in a particular direction. However, the data allows us to rule out such an explanation. Remember, that in Part II the subjects' behavior strongly resembles that one of the subjects in Charness and Rabin (2002) as well as in Chen and Li (2009) in those rounds that are independent of the Power Game (see Table 4). Thus, we argue that Part I did not alter the subjects in any obvious direction.

5.9. **Choice set attributes.** One might argue that changes in subjects' choices across Parts I and II are not due to differences in how much power player A has over B, but are due to the differences in the attributes of the choices sets (interval vs. fixed value) *per se*.

Two elements support that our design is one that indeed identifies preferences for power. First, the fact that subjects have more power over B players in Part I than in Part II conforms to the notion of power that has been discussed in the social psychology literature. For example, Keltner, Gruenfeld and Anderson (2003) emphasize that power is the *relative* capacity to modify others' outcomes and that it should be characterized not in absolute terms but as falling on a continuum. Second, our vignette studies further show that subjects generally view individuals who can choose a subordinate's compensation from an interval as having more power than those who cannot.

However, several studies have demonstrated that removing or adding (presumed irrelevant) alternatives can affect decisions one makes over others. For example, List (2007) and Bardsley (2008) show that including the option to "take" in a dictator game significantly reduced giving, even though the option to not give was already in the choice set. While, in the subsequent literature, such changes in behavior are mostly explained in terms of intention-based social preferences,[31] what our work shows is that some of these subjects' choices may have been in fact motivated by power.

5.10. **Warm glow.** One may also wonder whether elements such as "warm glow" can explain our results. In Andreoni (1990), the author defines warm glow in a public good context and shows that an individual may contribute to a public good not because she cares about the public good per se, but because giving makes her feel good about herself. This brings about the possibility that we mis-identify our subjects' motives when making decisions. In Andreoni (1990) or papers that test his theory (for example, Andreoni (1995), Crumpler and Grossman (2008)), warm glow is directly related to the level of contribution. In our experiment, this means that subjects who experience warm glow would experience the same level of it for a given x_B^*. Thus, strictly speaking, an individual who makes decisions because she is motivated by warm glow should make the same decisions in both Parts. In our classification these subjects have social preferences. As such, they may in fact be motivated by warm glow.

Moving away from a strict interpretation of Andreoni's concept, one may wonder whether the level of warm glow increases with the size of the choice set. In Part II the intensity of warm glow could be lessened by the fact that there are only two fixed alternatives for B's payoff. If this motivated the decision to pay in Part I and not pay in Part II, as Power+ subjects do, we should see that they give the maximum allowable in Part I, since it it costless to do so once the price to choose for B has been paid. However, Power+ subjects' behavior is at odds with this prediction. Thus, warm glow, whether in a strict sense or not, cannot explain the behavior of Power+ subjects.

6. Conclusion

In this paper we introduce a new game, the Power Game, and use it to identify individuals who have preferences for power—the ability to determine payoffs of others—without confounding other elements that may exist in the presence of power. Our work is the first to identify such preferences. We find that more than a quarter of the population values power *per se*, beyond its instrumental value. We show that these preferences for power are different than, and cannot be explained by, outcome- or intentions-based social preferences. Moreover, we show that our results are not due to specific design choices or other factors, such as mistakes, confusion, warm glow, experimenter demand or time trends. The vast majority of subjects who value power, do so in the absence of social preferences: they attach little value to other people's

[31]For example, as desire to signal one's kindness (Cappelen et al. (2013)), desire to follow social norms (Krupka and Weber (2013)), or preference for not taking to giving (Korenok, Millner and Razzolini (2014)).

outcomes and instead enjoy being the ones to choose those outcomes. Given that power-hungry subjects choose efficient allocations significantly less often than subjects who have social preferences, our results imply that social welfare is likely to decrease when individuals with power preferences are the ones allocating resources.

Results from our additional treatments reveal several interesting mechanisms that underlay preferences for power. For a given choice space, individuals value power over another individual more than over an organization. Additionally, power is related to the ability to directly choose payoffs of others, as opposed to simply influencing them. Further, individuals also value power less when their choice space is restricted. Thus, the value of power strongly depends on how flexible the decision-maker can be when making her choices, as well as how impactful her choices are.

Up until now, desires for power, control and autonomy had not been disentangled. We show that a large fraction of individuals seek power even if it does not grant them more control or autonomy from others. Thus, while these motives may still be present, a desire for control and non-interference from others are not the only motivation to climbing the ladder to the top. Further, since power is directed towards others, as opposed to control or autonomy, power preferences may have large and possibly problematic societal implications. Consequently, our findings provide strong reasons for incorporating preferences for power in the study of political systems, labor contracts and work relationships.

The millennia-old adage "the measure of a man is what he does with power," attributed to Plato, remains very relevant in our modern era and in the context studied in this paper. For example, when an individual runs for a political office, is it because he or she wishes to improve social welfare, or is it part of a quest for power? Similarly, one may question why a CEO proposes an acquisition. Is it to increase shareholder value, or is it for the purposes of empire building?

Understanding people's preferences as they relate to others is a challenging task and great strides have been made in this direction. Up until now, the study of outcome-based and intention-based social preferences has been the main focus of such work. We show that a substantial fraction of the population enjoys the *process* of choosing payoffs of others, without receiving additional utility from the actual payoff itself. Such individuals are willing to give up substantial amounts of money in order to engage in this process.

Much remains to be explored. For example, while in our anonymous experimental setting, subjects mostly demonstrated positive preferences for power, in other settings their willingness to exercise power may be diminished in the presence of external factors, e.g. blame by others, the possibility of retribution, etc. In such settings, other aspects of power may become more relevant, such as the ability to steer and restrict the actions of others, as opposed to the ability to choose their payoffs directly. Another avenue for future research is to explore how individual characteristics correlate with preferences for power, e.g. gender, education, and cultural background. Additionally, preferences for power could be incorporated in, and augment models of, optimal contract design, optimal organization structure, or even intentions-based or other non-outcome based preference models. We hope that our work will serve as a catalyst for new empirical and theoretical research in this area.

REFERENCES

Abbink, Klaus, and Abdolkarim Sadrieh. 2009. "The pleasure of being nasty." Economics Letters, 105(3): 306–308.

Aghion, Philippe, and Patrick Bolton. 1992. "An incomplete contracts approach to financial contracting." The Review of Economic Studies, 59(3): 473–494.

Agranov, Marina, and Pietro Ortoleva. 2017. "Stochastic choice and preferences for randomization." Journal of Political Economy, 125(1): 40–68.

Andreoni, James. 1990. "Impure altruism and donations to public goods: A theory of warm-glow giving." The Economic Journal, 100(401): 464–477.

Andreoni, James. 1995. "Warm-glow versus cold-prickle: The effects of positive and negative framing on cooperation in experiments." The Quarterly Journal of Economics, 110(1): 1–21.

Barclay, Michael J, and Clifford G Holderness. 1989. "Private benefits from control of public corporations." Journal of Financial Economics, 25(2): 371–395.

Bardsley, Nicholas. 2008. "Dictator game giving: Altruism or artefact?" Experimental Economics, 11(2): 122–133.

Bartling, Björn, and Urs Fischbacher. 2011. "Shifting the blame: On delegation and responsibility." The Review of Economic Studies, 79(1): 67–87.

Bartling, Björn, Ernst Fehr, and Holger Herz. 2014. "The intrinsic value of decision rights." Econometrica, 2005–2039.

Battigalli, Pierpaolo, and Martin Dufwenberg. 2007. "Guilt in games." The American Economic Review, 97(2): 170–176.

Battigalli, Pierpaolo, and Martin Dufwenberg. 2009. "Dynamic psychological games." Journal of Economic Theory, 144(1): 1–35.

Bendahan, Samuel, Christian Zehnder, François P Pralong, and John Antonakis. 2015. "Leader corruption depends on power and testosterone." The Leadership Quarterly, 26(2): 101–122.

Blount, Sally. 1995. "When social outcomes aren't fair: The effect of causal attributions on preferences." Organizational Behavior and Human Decision Processes, 63(2): 131–144.

Bolton, Gary E, and Axel Ockenfels. 2000. "ERC: A theory of equity, reciprocity, and competition." The American Economic Review, 166–193.

Bown, Nicola J, Daniel Read, and Barbara Summers. 2003. "The lure of choice." Journal of Behavioral Decision Making, 16(4): 297–308.

Brandts, Jordi, and Gary Charness. 2011. "The strategy versus the direct-response method: a first survey of experimental comparisons." Experimental Economics, 14(3): 375–398.

Brosig-Koch, Jeannette, Thomas Riechmann, and Joachim Weimann. 2017. "The dynamics of behavior in modified dictator games." PLOS ONE, 12(4): e0176199.

Cappelen, Alexander W, Ulrik H Nielsen, Erik Ø Sørensen, Bertil Tungodden, and Jean-Robert Tyran. 2013. "Give and take in dictator games." Economics Letters, 118(2): 280–283.

Charness, Gary. 2004. "Attribution and reciprocity in an experimental labor market." Journal of Labor Economics, 22(3): 665–688.

Charness, Gary, and David I Levine. 2007. "Intention and stochastic outcomes: An experimental study." The Economic Journal, 117(522): 1051–1072.

Charness, Gary, and Martin Dufwenberg. 2006. "Promises and partnership." Econometrica, 74(6): 1579–1601.

Charness, Gary, and Matthew Rabin. 2002. "Understanding social preferences with simple tests." The Quarterly Journal of Economics, 117(3): 817–869.

Charness, Gary, David Masclet, and Marie Claire Villeval. 2014. "The dark side of competition for status." Management Science, 60(1): 38–55.

Chassang, Sylvain, and Christian Zehnder. 2019. "Secure Survey Design in Organizations: Theory and Experiments." Working paper.

Chen, Yan, and Sherry Xin Li. 2009. "Group identity and social preferences." The American Economic Review, 99(1): 431–457.

Coffman, Lucas C. 2011. "Intermediation reduces punishment (and reward)." American Economic Journal: Microeconomics, 3(4): 77–106.

Cox, James C, Daniel Friedman, and Steven Gjerstad. 2007. "A tractable model of reciprocity and fairness." Games and Economic Behavior, 59(1): 17–45.

Cox, James C, Daniel Friedman, and Vjollca Sadiraj. 2008. "Revealed altruism." Econometrica, 76(1): 31–69.

Crumpler, Heidi, and Philip J Grossman. 2008. "An experimental test of warm glow giving." Journal of Public Economics, 92(5): 1011–1021.

Dahya, Jay, Orlin Dimitrov, and John J McConnell. 2008. "Dominant shareholders, corporate boards, and corporate value: A cross-country analysis." Journal of Financial Economics, 87(1): 73–100.

Demsetz, Harold, and Kenneth Lehn. 1985. "The structure of corporate ownership: Causes and consequences." Journal of Political Economy, 93(6): 1155–1177.

De Quidt, Jonathan, Johannes Haushofer, and Christopher Roth. 2018. "Measuring and bounding experimenter demand." The American Economic Review, 108(11): 3266–3302.

de Quidt, Jonathan, Lise Vesterlund, and Alistair J Wilson. 2018. "Experimenter demand effects." Handbook of Research Methods and Applications in Experimental Economics, forthcoming.

Dessein, Wouter, and Richard Holden. 2019. "Organizations with power-hungry agents." CEPR Discussion Paper No. DP13526.

Doidge, Craig, G Andrew Karolyi, Karl V Lins, Darius P Miller, and René M Stulz. 2009. "Private benefits of control, ownership, and the cross-listing decision." The Journal of Finance, 64(1): 425–466.

Dufwenberg, Martin, and Georg Kirchsteiger. 2000. "Reciprocity and wage undercutting." European Economic Review, 44(4-6): 1069–1078.

Dufwenberg, Martin, and Georg Kirchsteiger. 2004. "A theory of sequential reciprocity." Games and Economic Behavior, 47(2): 268–298.

Dyck, Alexander, and Luigi Zingales. 2004. "Private benefits of control: An international comparison." The Journal of Finance, 59(2): 537–600.

Ederer, Florian, and Alexander Stremitzer. 2017. "Promises and expectations." Games and Economic Behavior, 106: 161–178.

Ellingsen, Tore, Magnus Johannesson, Sigve Tjøtta, and Gaute Torsvik. 2010. "Testing guilt aversion." Games and Economic Behavior, 68(1): 95–107.

Emerson, Richard M. 1962. "Power-dependence relations." The American Sociological Review, 31–41.

Falk, Armin, and Urs Fischbacher. 2006. "A theory of reciprocity." Games and Economic Behavior, 54(2): 293–315.

Falk, Armin, Ernst Fehr, and Urs Fischbacher. 2003. "On the nature of fair behavior." Economic Inquiry, 41(1): 20–26.

Falk, Armin, Ernst Fehr, and Urs Fischbacher. 2008. "Testing theories of fairness – Intentions matter." Games and Economic Behavior, 62(1): 287–303.

Fehr, Ernst, and Klaus M Schmidt. 1999. "A theory of fairness, competition, and cooperation." The Quarterly Journal of Economics, 114(3): 817–868.

Fehr, Ernst, Holger Herz, and Tom Wilkening. 2013. "The lure of authority: Motivation and incentive effects of power." The American Economic Review, 103(4): 1325–1359.

Ferreira, Joao V, Nobuyuki Hanaki, and Benoît Tarroux. 2017. "On the roots of the intrinsic value of decision rights: Evidence from France and Japan." Center for Research in Economics and Management (CREM), University of Rennes 1, University of Caen and CNRS.

Fershtman, Chaim, and Uri Gneezy. 2001. "Strategic delegation: An experiment." RAND Journal of Economics, 352–368.

Fischbacher, Urs. 2007. "z-Tree: Zurich toolbox for ready-made economic experiments." Experimental Economics, 10(2): 171–178.

Fiske, Susan T. 1993. "Controlling other people: The impact of power on stereotyping." American Psychologist, 48(6): 621.

Fiske, Susan T, and Eric Dépret. 1996. "Control, interdependence and power: Understanding social cognition in its social context." European Review of Social Psychology, 7(1): 31–61.

Geanakoplos, John, David Pearce, and Ennio Stacchetti. 1989. "Psychological games and sequential rationality." Games and Economic Behavior, 1(1): 60–79.

Grossman, Sanford J, and Oliver D Hart. 1986. "The costs and benefits of ownership: A theory of vertical and lateral integration." Journal of Political Economy, 94(4): 691–719.

Gul, Faruk, and Wolfgang Pesendorfer. 2016. "Interdependent preference models as a theory of intentions." Journal of Economic Theory, 165: 179–208.

Hart, Oliver D, and John Moore. 1995. "Debt and seniority: An analysis of the role of hard claims in constraining management." The American Economic Review, 85(3): 567–585.

Hart, Oliver D, and John Moore. 2005. "On the design of hierarchies: Coordination versus specialization." Journal of Political Economy, 113(4): 675–702.

Jensen, Michael C, and William H Meckling. 1976. "Theory of the firm: Managerial behavior, agency costs and ownership structure." Journal of Financial Economics, 3(4): 305–360.

Keltner, Dacher, Deborah H Gruenfeld, and Cameron Anderson. 2003. "Power, approach, and inhibition." Psychological Review, 110(2): 265.

Korenok, Oleg, Edward L Millner, and Laura Razzolini. 2014. "Taking, giving, and impure altruism in dictator games." Experimental Economics, 17(3): 488–500.

Krupka, Erin L, and Roberto A Weber. 2013. "Identifying social norms using coordination games: Why does dictator game sharing vary?" Journal of the European Economic Association, 11(3): 495–524.

Lazear, Edward P, Ulrike Malmendier, and Roberto A Weber. 2012. "Sorting in experiments with application to social preferences." American Economic Journal: Applied Economics, 4(1): 136–63.

Levine, David K. 1998. "Modeling altruism and spitefulness in experiments." Review of Economic Dynamics, 1(3): 593–622.

List, John A. 2007. "On the interpretation of giving in dictator games." Journal of Political Economy, 115(3): 482–493.

List, John A, Azeem M Shaikh, and Yang Xu. 2016. "Multiple hypothesis testing in experimental economics." Experimental Economics, 1–21.

Magee, Joe C, and Adam D Galinsky. 2008. "8 social hierarchy: The self-reinforcing nature of power and status." Academy of Management Annals, 2(1): 351–398.

McCabe, Kevin A, Stephen J Rassenti, and Vernon L Smith. 1998. "Reciprocity, trust, and payoff privacy in extensive form bargaining." Games and Economic Behavior, 24(1-2): 10–24.

Neri, Claudia, and Hendrik Rommeswinkel. 2017. "Decision rights: Freedom, power, and interference." Working paper.

Orhun, A Yeşim. 2018. "Perceived motives and reciprocity." Games and Economic Behavior, 109: 436–451.

Owens, David, Zachary Grossman, and Ryan Fackler. 2014. "The control premium: A preference for payoff autonomy." American Economic Journal: Microeconomics, 6(4): 138–161.

Rabin, Matthew. 1993. "Incorporating fairness into game theory and economics." The American Economic Review, 1281–1302.

Rand, David G, Drew Fudenberg, and Anna Dreber. 2015. "It's the thought that counts: The role of intentions in noisy repeated games." Journal of Economic Behavior & Organization, 116: 481–499.

Sebald, Alexander. 2010. "Attribution and reciprocity." Games and Economic Behavior, 68(1): 339–352.

Segal, Uzi, and Joel Sobel. 2007. "Tit for tat: Foundations of preferences for reciprocity in strategic settings." Journal of Economic Theory, 136(1): 197–216.

Selten, Reinhard. 1967. "Die Strategiemethode zur Erforschung des engeschraänkt rationalen Verhaltens im Rahmen eines Oligopolexperiments." In Beiträge zuer Experimentellen Wirtschaftsforschung. , ed. H. Sauermann, 136–168. Tübingen:J. C. B. Mohr.

Sloof, Randolph, and Ferdinand A von Siemens. 2017. "Illusion of control and the pursuit of authority." Experimental Economics, 20(3): 556–573.

Sobel, Joel. 2005. "Interdependent preferences and reciprocity." Journal of Economic Literature, 43(2): 392–436.

Tost, Leigh Plunkett. 2015. "When, why, and how do powerholders "feel the power"? Examining the links between structural and psychological power and reviving the connection between power and responsibility." Research in Organizational Behavior, 35: 29–56.

Toussaert, Séverine. 2017. "Intention-based reciprocity and signaling of intentions." Journal of Economic Behavior & Organization, 137: 132–144.

Vanberg, Christoph. 2008. "Why do people keep their promises? An experimental test of two explanations." Econometrica, 76(6): 1467–1480.

Zizzo, Daniel John, and Andrew J Oswald. 2001. "Are people willing to pay to reduce others' incomes?" Annales d'Economie et de Statistique, 39–65.

FOR ONLINE PUBLICATION

APPENDIX A.

Below we present the instructions that the subjects received in our experiment.

INSTRUCTIONS

This study is in 4 Parts. In each Part you will participate in a number of Rounds. Only one Part will be chosen for payment and only one Round in that Part will count towards payment. In addition to what you will earn in the study, you will get a 7-dollar participation fee if you complete the study. After you have completed the 4 Parts of this study, you will be able to see your total earnings on your screen. In order to preserve the anonymity of everyone in this room, you will be given no other information.

Importantly, nothing you do in any Round or Part of this study can reduce or increase your potential payment in any other Round or Part of this study. In addition, since only one randomly chosen Round in one randomly chosen Part will be chosen for payment, it is in your best interest to treat each Part and each Round as if it was the only one that mattered for payment.

We will now hand out the instructions for Part 1 of the study. We will give you the instructions for Part 2 of the study once you have completed Part 1, for Part 3 after you have completed Part 2, and for Part 4 after you have completed Part 3.

The List of Choices in Part 1

Decision Problem	Option 1		Option 2
1	Fixed amount of $0	Versus	Let the computer choose between $0 and $16.30
2	Fixed amount of $3.10	Versus	Let the computer choose between $0 and $16.30
3	Fixed amount of $6.60	Versus	Let the computer choose between $0 and $16.30
4	Fixed amount of $12.30	Versus	Let the computer choose between $0 and $16.30
5	Fixed amount of $16.30	Versus	Let the computer choose between $0 and $16.30

PART 1

In this Part of the study, you will make decisions over the course of 5 Rounds. In each Round, you will be asked to choose between two options that determine your payoff. Below we list the exact 5 decision problems that you will all face.

These decision problems may appear in different order on your screen. In addition, for any given decision problem, which option appears on the left or the right of your screen may also differ from the examples above.

In each of these Rounds you will have the choice between a fixed amount or letting the computer choose your outcome. If you let the computer choose, the computer will determine your outcome by dividing the $0 to $16.30 range in 5 cent increments and randomly choosing one of those numbers, with each number having an equal chance of being chosen. In other words, the computer will randomly choose a number among $0, 5 cents, 10 cents, 15 cents ... etc. up until $16.30. Note that any number on that interval that is a multiple of 5 cents, that is $0.00, $0.05, $0.10, ... $16.20, $16.25, $16.30, is equally likely to be chosen.

If this Part of the study is chosen to count for payment, then one of the 5 Rounds will be randomly chosen to count for payment. If in that Round, you choose the fixed amount, then you will receive that fixed amount. If in that Round, you choose to let the computer choose, then your payoff will be randomly determined on the interval between $0 and $16.30 (both included) in intervals of 5 cents.

PREAMBLE TO THE REST OF THE STUDY

Before we continue this study, you will be assigned a Type. You will be a Type A Player or a Type B Player. Your Type will remained fixed throughout the rest of this entire study and will be relevant for the rest of this study (Parts 2, 3 and 4).

In each of Parts 2, 3 and 4, at the start of each Round, each Type A Player will be randomly rematched with a Type B Player. You will not know who you are matched with. In this study, only Type A players make decisions that matter for payment, and these decisions affect the payoff of both the Type A Player and the Type B Player he/she is matched with.

Even though your Type will remain fixed for the rest of this study, you will not know which Type of Player you are. Since you do not know which Type of Player you are assigned to be, and since only Type A Players make decisions that matter for payment, we will ask everyone to make decisions as if they were Type A players.

Please note that your Type will remain fixed for the rest of this study and at no point will you change roles. Your "true" Types have already been determined by the computer, and your decisions when acting as Player A CANNOT affect you or anyone else in this room if your "true" Type turns out to be Type B. In other words, if it turns out you are a Type B Player, no decision you make here can affect anyone's payoff, including your own. If it turns out your "true" Type is A, there is nothing that anyone else can do that will affect your payoff, and your decisions affect both your payoff and the payoff of the Type B Player you are matched with. Therefore, when making decisions, you should act as Player A. Further, since only "true" Type A Players make decisions that matter for payment in this study, in the remainder of the instructions we will assume you are a Type A Player.

PART 2

In this Part of the study, you will make decisions over the course of 11 Rounds. As a Type A Player, in each Round, you can choose to pay a certain amount of money to obtain the right to choose the payoff of the Type B Player you are matched with. The price of that right will vary from Round to Round.

If you choose to pay this amount, you will be given $12.30 and the price will be subtracted from the $12.30 you have. If you obtain the right to choose the payoff for the Type B Player you are matched with, you can choose any number between $0 and $16.30, both included, by increments of 5 cents.

If you choose to not pay that amount of money, you do not obtain the right to choose Type B's payoff and both you and the Type B Player you are matched with earn $12.30.

Here is an example of such a choice you can encounter in one of the Rounds (the choices you face may be different and will vary from Round to Round). Suppose the price for obtaining the right to choose the payoff for Player B is $1.

- If you choose to pay $1 then you can choose Player B's payoff.
  - Suppose you choose $3 as Player B's payoff. In this case, if this Round is chosen for payment, you will earn $12.30 - $1 = $11.30 and the Type B Player you are matched with will earn $3.
  - If instead you choose $11.55 as Player B's payoff, and if this Round is chosen for payment, you will still earn $12.30 - $1 = $11.30 and the Type B Player you are matched with will earn $11.55.
- If you choose not to pay $1 then you cannot choose Player B's payoff. In this case, if this Round is chosen for payment, you will earn $12.30 and the Type B Player you are matched with will earn $12.30.

The screen you face would look like the following:

In this Round, **the price for the right to choose the payoff for Player B is $1.00.**
If you choose to pay, your own payoff will be $12.30-$1 = $11.30 and you will obtain the right to choose B’s payoff.
If you choose not to pay, both you and B will have a payoff of $12.30.
Do you wish to choose Player B’s payoff?

Your Payoff: $12.30 B’s Payoff: $12.30	Your Payoff: $11.30 B’s Payoff: You obtain the right to choose B’s payoff and can choose between $0 and $16.30 (both included)
Don’t choose B’s payoff	Choose B’s payoff

Note that if you choose to pay and obtain the right to choose B’s payoff, then you will face a new screen where you will enter your choice for B. In order to preserve the anonymity of everyone’s choices, even if you choose to not pay and you do not obtain the right to choose B’s payoff, you will still face a new screen where you will be asked to type in between 1 and 5 characters (letters, numbers, and special characters are all allowed), though in this case nothing you type will have any impact on anyone’s payoff.

Remember that you will not change roles in this study. So as a Type A Player, your payoff will never be determined by someone else in this room. Also remember that only one Part of the study will be chosen to count for payment. If this Part is chosen to count, only *one* Round will matter for payment. So it is in your best interest to treat each Round as if it were the one that mattered for payment.

Before we start the 11 Rounds, we will show you three screens so that you can familiarize yourselves with the interface. The first screen will be an example of a screen in which you have to decide whether to obtain the right to choose B’s payoff or not. The second screen will be what you would see if you did not pay for the right to choose Type B’s payoff. The third screen will be what you would see if you did pay for the right to choose Type B’s payoff. These "practice" screens do not count towards payment.

PART 3

Before we describe Part 3, recall that if your "true" Type is B, nothing you do here will have any impact on anyone's payoff, including your own. Therefore you should act as a Type A Player and in what follows we will assume that you are indeed a Type A Player.

In each Round of this Part of the study, you will be asked to choose between two options that determine payoffs for both you and the Type B Player you are matched with. Here is an example of such a choice you can encounter in one of the Rounds (the choices you face may be different and will vary from Round to Round):

Your Payoff: $6.15 B's Payoff: $12.30	Your Payoff: $11.30 B's Payoff: $9.00
Choose this option	Choose this option

Here is an example, assuming that Part 2 and this Round was chosen for payment. If you choose the pair on the left, you will earn $6.15 and the Type B Player you are matched with will earn $12.30. If choose the pair on the right, you will earn $11.30 and the Type B Player you are matched with will earn $9.00.

You will play 22 Rounds of this Game.

Remember that you will not change roles in this study. So as a Type A Player, your payoff will never be determined by someone else in this room. Also remember that only one Part of the study will be chosen to count for payment. If this Part is chosen to count, only *one* Round will matter for payment. So it is in your best interest to treat each Round as if it were the one that mattered for payment.

PART 4

Before we describe Part 4, recall that if your "true" Type is B, nothing you do here will have any impact on anyone's payoff, including your own. Therefore you should act as a Type A Player and in what follows we will assume that you are indeed a Type A Player.

The task in this last Part of the study is the same as the task you faced in Part 2. The instructions below serve to remind you of that task.

In this Part of the study, you will make decisions over the course of 11 Rounds. As a Type A Player, in each Round, you can choose to pay a certain amount of money to obtain the right to choose the payoff of the Type B Player you are matched with. The price of that right will vary from Round to Round.

If you choose to pay this amount, you will be given $12.30 and the price will be subtracted from the $12.30 you have. If you obtain the right to choose the payoff for the Type B Player you are matched with, you can choose any number between $0 and $16.30, both included, by increments of 5 cents.

If you choose to not pay that amount of money, you do not obtain the right to choose Type B's payoff and both you and the Type B Player you are matched with earn $12.30.

Here is an example of such a choice you can encounter in one of the Rounds (the choices you face may be different and will vary from Round to Round). Suppose the price for obtaining the right to choose the payoff for Player B is $1.

- If you choose to pay $1 then you can choose Player B's payoff.
  - Suppose you choose $3 as Player B's payoff. In this case, if this Round is chosen for payment, you will earn $12.30 - $1 = $11.30 and the Type B Player you are matched with will earn $3.
  - If instead you choose $11.55 as Player B's payoff, and if this Round is chosen for payment, you will still earn $12.30 - $1 = $11.30 and the Type B Player you are matched with will earn $11.55.
- If you choose not to pay $1 then you cannot choose Player B's payoff. In this case, if this Round is chosen for payment, you will earn $12.30 and the Type B Player you are matched with will earn $12.30.

The screen you face would look like the following:

In this Round, **the price for the right to choose the payoff for Player B is $1.00.**
If you choose to pay, your own payoff will be $12.30-$1 = $11.30 and you will obtain the right to choose B's payoff.
If you choose not to pay, both you and B will have a payoff of $12.30.
Do you wish to choose Player B's payoff?

Your Payoff: $12.30 B's Payoff: $12.30	Your Payoff: $11.30 B's Payoff: You obtain the right to choose B's payoff and can choose between $0 and $16.30 (both included)
Don't choose B's payoff	Choose B's payoff

Note that if you choose to pay and obtain the right to choose B's payoff, then you will face a new screen where you will enter your choice for B. In order to preserve the anonymity of everyone's choices, even if you choose to not pay and you do not obtain the right to choose B's payoff, you will still face a new screen where you will be asked to type in between 1 and 5 characters (letters, numbers, and special characters are all allowed), though in this case nothing you type will have any impact on anyone's payoff.

Remember that you will not change roles in this study. So as a Type A Player, your payoff will never be determined by someone else in this room. Also remember that only one Part of the study will be chosen to count for payment. If this final Part is chosen to count, only *one* Round will matter for payment. So it is in your best interest to treat each Round as if it were the one that mattered for payment.

Appendix B.

Below we present the screens subjects saw in different tasks of our experiment.

Figure B1. Lottery task screen faced by the subjects, Task 1 of the experiment.

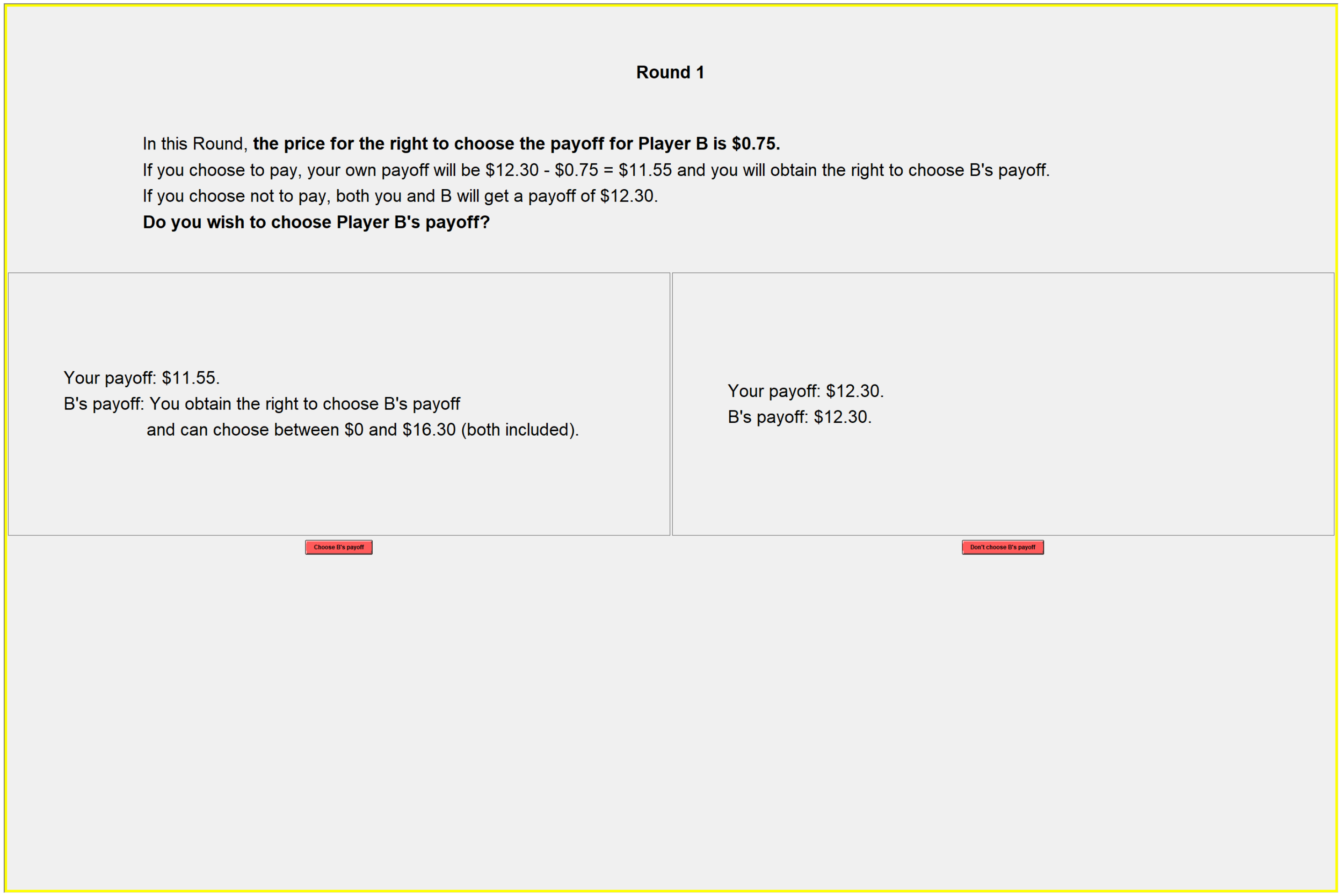

FIGURE B2. First-stage screen with one of the decisions faced by the subjects in Part I of the Power Game, Tasks 2 and 4 of the experiment.

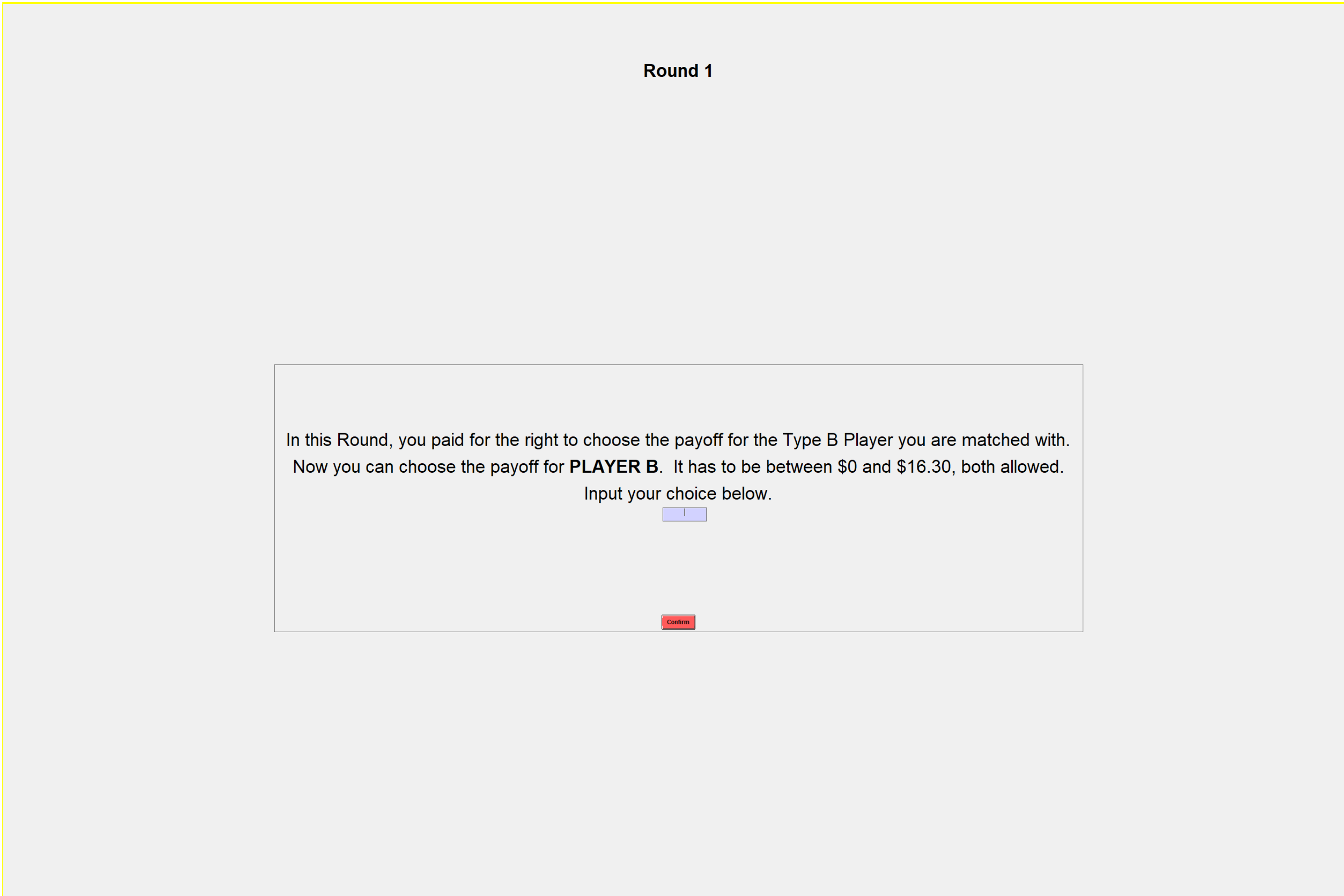

FIGURE B3. Second-stage screen faced by the subjects in Part I of the Power Game if they chose to pay to choose *B*'s payoff, Tasks 2 and 4of the experiment.

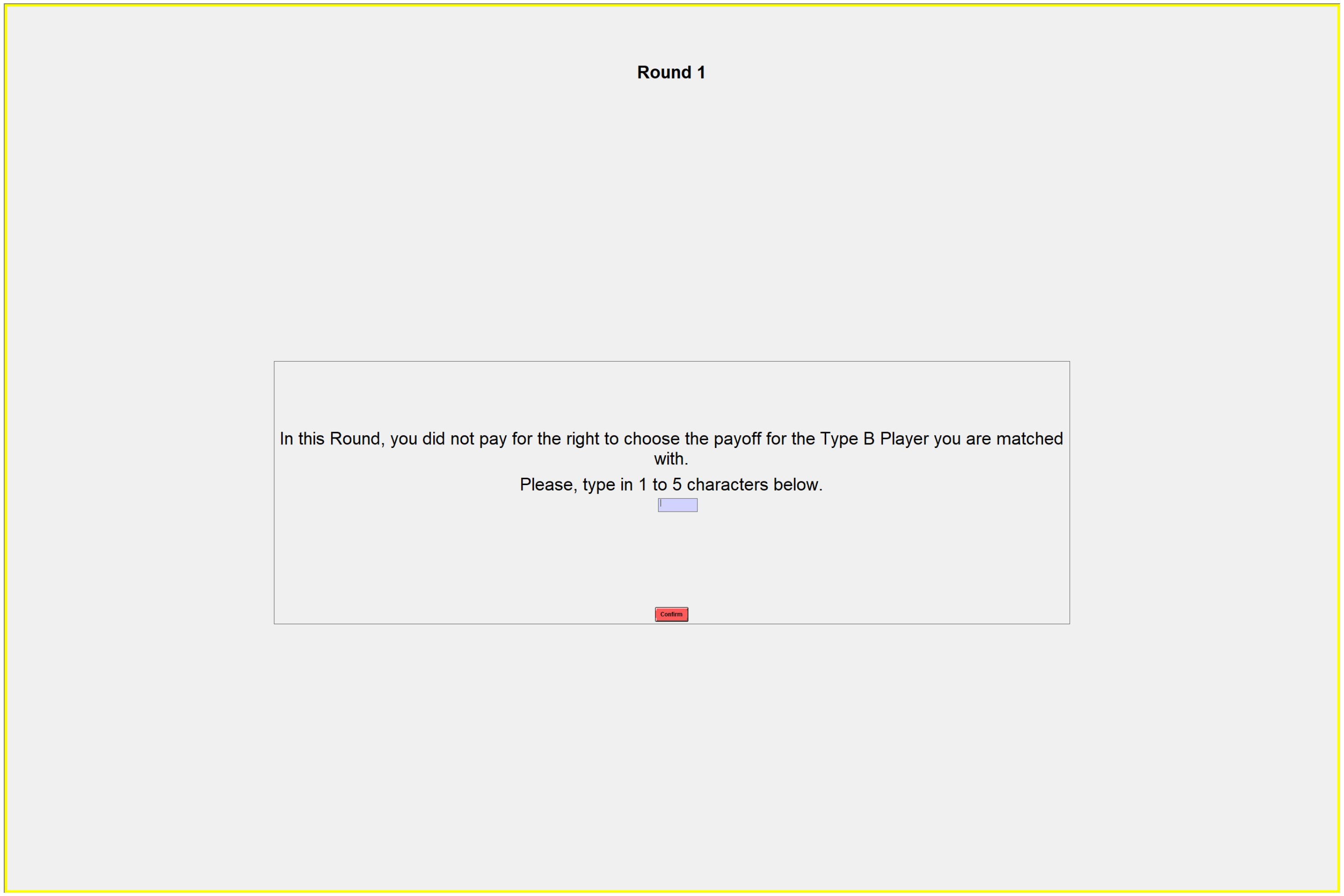

FIGURE B4. Second-stage screen faced by the subjects in Part I of the Power Game if they chose not to pay to choose B's payoff, Tasks 2 and 4 of the experiment.

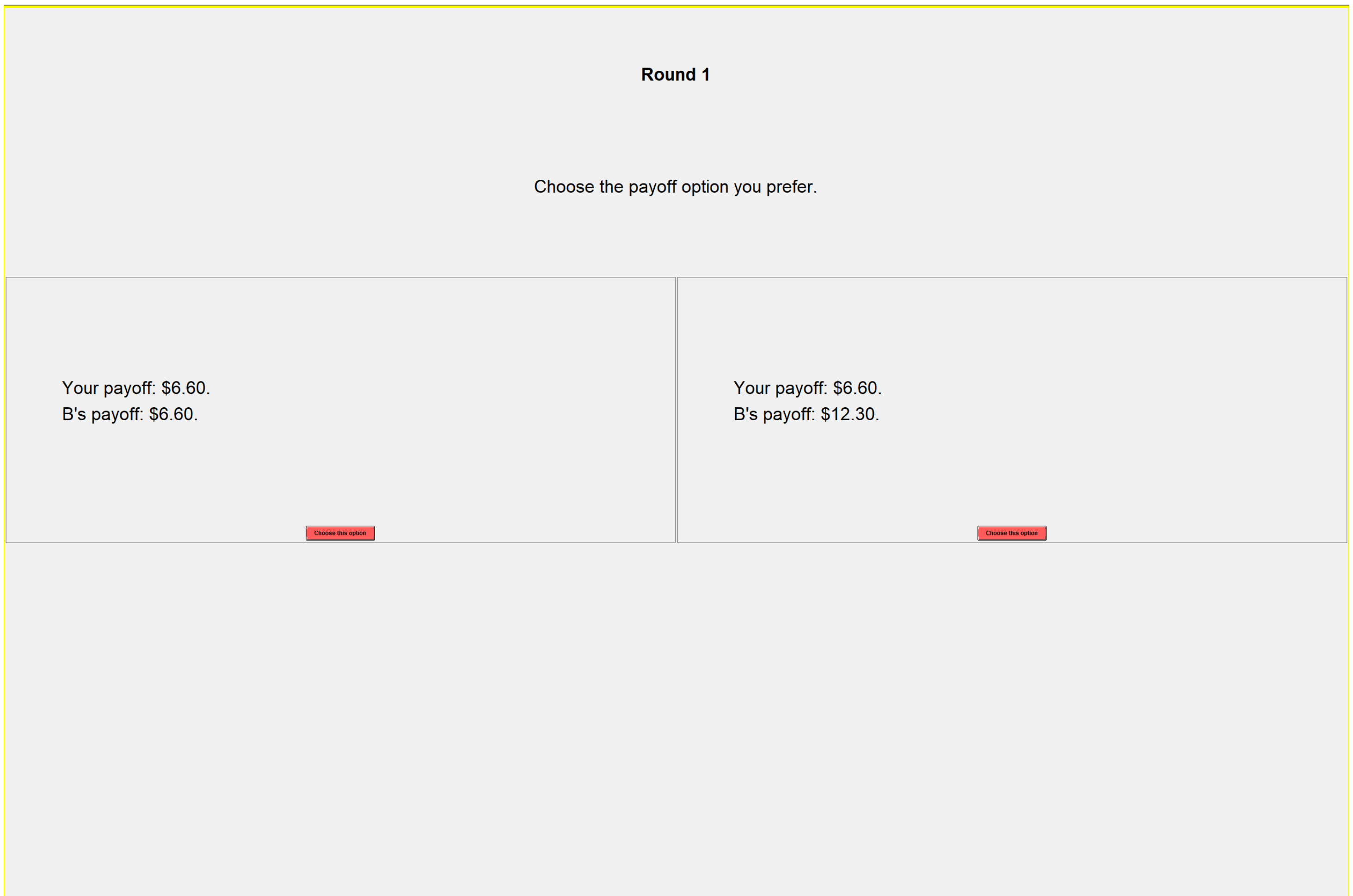

FIGURE B5. Screen with one of the decisions faced by the subjects in Part II of the Power Game, Task 3 of the experiment.

APPENDIX C.

Below we list the questions that we asked all subjects to answer at the end of the experiment.

I. Demographics
   (a) Please, enter your age:
   (b) Which gender do you identify with?
      - Male
      - Female
      - Other
   (c) How many years of university education have you received?
   (d) Are you a native English speaker?
   (e) Which faculty best describes your field of study?
      - Science
      - Social science
      - Arts
      - Engineering
      - Business
      - Other
   (f) What is the ZIP code of the place where you grew up?
   (g) What is the highest degree your mother has?
      - Less than high school
      - High school or equivalent
      - Some college
      - College
      - More than college
      - Other
      - Not sure
   (h) What is the highest degree your father has?
      - Less than high school
      - High school or equivalent
      - Some college
      - College
      - More than college
      - Other
      - Not sure
   (i) Have you ever participated in similar experiments before?

II. Understanding of the game
   (a) Was anything confusing?
   (b) What do you think the experiment was about?
   (c) What motivated your choices in Part 2?
   (d) What motivated your choices in Part 3?
   (e) What motivated your choices in Part 4?

APPENDIX D.

We start by specifying player A's utility function in a general form: $\text{U}_A = \text{U}(x_A, x_B, \lambda)$, where x_A is A's own payoff, x_B is B's payoff, and λ is a parameter indicating the amount of power that A has over B's payoff. Without loss of generality, we normalize λ to zero when B's potential payoff is pre-specified, i.e. when A receives no additional power from her choice, as is the case in all Part II rounds as well as if she doesn't pay in Part I. Similarly, we impose $\lambda = 1$ when A chooses B's payoff from the $[0, E_B]$ interval. We make the following assumptions about A's utility function.

Assumption 1. $\text{U}(x_A, x_B, \lambda)$ *is continuous in all three arguments.*

Assumption 2. *For all x_B and all λ, $\text{U}(x_A, x_B, \lambda)$ is strictly increasing in x_A. For power-hungry (power-averse) players, $\text{U}(x_A, x_B, \lambda)$ is strictly increasing (decreasing) in λ. For those indifferent to power, $\text{U}(x_A, x_B, \lambda) = \text{U}(x_A, x_B, \lambda')$ for all λ, λ'.*

Assumption 3. *If $x_A > x'_A$ and $\text{U}(x_A, x_B, \lambda) = \text{U}(x'_A, x'_B, \lambda')$, then a player chooses the (x_A, x_B, λ) option.*

Assumption 1 simply states that player A's utility function is continuous for all values of its arguments x_A, x_B, and λ. In Assumption 2 we impose that all else equal, A's utility function is strictly increasing in her own payoff and strictly monotonic in power. Assumption 3 states that if a player is indifferent between two options, she always chooses the one that gives her the highest monetary payoff. In particular, if a player pays p, this represents a strict preference ordering. We include this assumption for convenience and it is innocuous vis-à-vis our results. Before we derive predictions about an individual's willingness to pay in Parts I and II of the Power Game we show that her demand function is well-behaved.

Lemma 1. *Player A has a well-behaved demand function in both Parts of the Power Game.*

Proof. Here we concentrate on behavior in Part I of the Power Game as the proof for Part II is similar and therefore omitted. First, we prove that if a subject pays p', then she pays for all $p < p'$. We then prove that if a subject does not pay p'', then she does not pay for any $p > p''$. These steps suffice to show that demand functions are well-behaved.

Suppose A pays p'. Then, we must have that $\text{U}(E_A - p', x^*_B, 1) > \text{U}(E_A, E_A, 0)$, where x^*_B is what A chooses for B. Since $\text{U}(x_A, x_B, \lambda)$ is strictly increasing in x_A, for any $p < p'$, we have that $\text{U}(E_A - p, x^*_B, 1) > \text{U}(E_A - p', x^*_B, 1) > \text{U}(E_A, E_A, 0)$. In other words, at price $p < p'$ player A is better off choosing x^*_B than not paying. Thus, for all $p < p'$, A pays and A's demand function is well-behaved for all $p < p'$.

Suppose A does not pay p''. Then, $\text{U}(E_A - p'', x^{**}_B, 1) < \text{U}(E_A, E_A, 0)$, where x^{**}_B is what A chooses for B at $p = p''$. By monotonicity of $\text{U}(x_A, x_B, \lambda)$ in x_A, for any $p > p''$ we have, for all x_B, that $\text{U}(E_A - p, x_B, 1) < \text{U}(E_A - p'', x^{**}_B, 1)$. Thus, for all $p > p''$, player A does not pay and A's demand function is well-behaved for all $p > p''$. □

Lemma 1 shows that in each Part of the Power Game players follow one of three paying behaviors: (1) a player never pays; (2) a player pays up until a cutoff price $\bar{p}_{\text{I}}$ but does not pay for any price above it; or (3) a player pays at all prices. Note that to guarantee a finite cutoff price, additional constraints are needed: for example, if A faces a budget constraint. Heretofore, we call player A's willingness to pay in Part I $\bar{p}_{\text{I}}$ and call A's willingness to pay in Part II $\bar{p}_{\text{II}}$. Note that there are no restriction as to the signs of $\bar{p}_{\text{I}}$ and $\bar{p}_{\text{II}}$.

We now turn to modeling the behavior of players in different preference classes. We begin with a player who has "standard" (completely selfish) preferences and who does not derive any utility from having power: $\text{U}(x_A, x_B, \lambda) > \text{U}(x'_A, x'_B, \lambda')$ for all $x_A > x'_A$, all x_B and x'_B, and all λ and λ'. Proposition 1 states that a player with standard preferences always chooses to maximize her own payoff and never pays positive prices in either Part of the Power Game.

Proposition 1. *For a player with standard preferences, $\bar{p}_I = \bar{p}_{II} = 0$.*

Proof. In Part I, if player A pays $p > 0$, then her payoff is equal to $E_A - p$. If she does not pay, her payoff is E_A. Since she only cares about her own payoff, she does not pay for any $p > 0$ and pays for all $p < 0$. Thus, $\bar{p}_{\text{I}} = 0$, and similarly $\bar{p}_{\text{II}} = 0$. □

Next, we consider the behavior of players with non-standard preferences, i.e. players with power preferences or social preferences or both. We start with players who like power but do not have social preferences. If a player enjoys being able to choose payoffs for others, her utility has to incorporate not only final payoffs, but also whether those payoffs are attained via increased power: for $\lambda > \lambda'$, $\text{U}(x_A, x_B, \lambda) > \text{U}(x_A, x_B, \lambda')$ for all x_A and x_B.

Proposition 2. *For a player with power preferences and no social preferences, $\bar{p}_I > 0$ and $\bar{p}_{II} = 0$.*

Proof. Since player A derives a positive utility from having power and does not have social preferences, then in Part I of the Power Game, for all x_B, $\text{U}(E_A, x_B, 1) > \text{U}(E_A, E_A, 0)$. By continuity and monotonicity of $\text{U}(x_A, x_B, \lambda)$ in x_A, there exists $p' > 0$ such that $\text{U}(E_A, x_B, 1) > \text{U}(E_A - p', x_B, 1) > \text{U}(E_A, E_A, 0)$, and A pays $p' > 0$. By Lemma 1, we conclude that $\bar{p}_{\text{I}} > 0$.

In Part II of the Power Game, paying does not lead to any increase in power for A, since in either case the potential payoff for B is fixed. Therefore, for any $p > 0$ and any x_B, x'_B, $\text{U}(E_A, x_B, 0) > \text{U}(E_A - p, x'_B, 0)$. Thus, $\bar{p}_{\text{II}} = 0$. □

Proposition 2 states that a player with power preferences and no social preferences is willing to pay positive prices in Part I of the Power Game. In Part II however she instead chooses the payoff-maximizing option and never pays a positive price.

Let us now consider a player who has social preferences and no power preferences. Such a player incorporates B's payoff in her utility in a non-trivial manner, but because she is indifferent towards power, her utility from a particular allocation (x_A, x_B) is not affected by how that allocation is obtained. In other words, this player's utility is the same whether she chooses a particular payoff for B from $[0, E_B]$ or whether it is exogenously given. For her, $\mathrm{U}(x_A, x_B, \lambda) = \mathrm{U}(x_A, x_B, \lambda')$ for any λ λ'.

We make an additional assumption: we assume that if A has social preferences, then her utility is highest at a single point in the $[0, E_B]$ interval. That is, for a given x_A there are no two different payoffs for B that give A that highest utility. Assumption 4 formalizes these notions.

Assumption 4. *For a player with social preferences, for any x_A and λ, there exists a unique x_B^* that maximizes her utility, that is* $\underset{x_B \in [0,E_B]}{\operatorname{argmax}} \mathrm{U}(x_A, x_B, \lambda) = \{x_B^*\}$.

Social preferences that satisfy our assumptions can be divided into two categories that are mutually exclusive and together comprise the entire set of social preferences:

(1) A maximizes her utility by choosing something other than E_A for player B when $p = 0$. That is, there exists $x_B^{\circledast} \neq E_A$ such that for all λ and λ', $\mathrm{U}(E_A, x_B^{\circledast}, \lambda) > \mathrm{U}(E_A, E_A, \lambda)$, i.e. $E_A \neq \underset{x_B \in [0,E_B]}{\operatorname{argmax}} \mathrm{U}(E_A, x_B, \lambda)$.[32]

(2) A maximizes her utility by choosing E_A for player B when $p = 0$. That is, for all λ and λ', $\mathrm{U}(E_A, E_A, \lambda) > \mathrm{U}(E_A, x_B, \lambda)$, for any $x_B \neq E_A$, i.e. $E_A = \underset{x_B \in [0,E_B]}{\operatorname{argmax}} \mathrm{U}(E_A, x_B, \lambda)$.[33]

In Proposition 3 and Corollary 1 we derive predictions regarding the paying behavior of players with social preferences and no power preferences in the Power Game.

Proposition 3. *For a player with social preferences and no power preferences, $\bar{p}_I > 0$ if and only if she does not choose E_A for player B when $p = 0$. In Part II, player A has the same paying behavior as in Part I: $\bar{p}_I = \bar{p}_{II}$.*

Proof. Let us start by showing that if A does not choose E_A for B when $p = 0$, then there exits some positive price $p' > 0$ such that A pays p'.

Let $x_B^{\circledast}$ be what A chooses for B when $p = 0$, i.e. $x_B^{\circledast} = \underset{x_B \in [0,E_B]}{\operatorname{argmax}} \mathrm{U}(E_A, x_B, 1)$, which is a singleton by Assumption 4. Since $x_B^{\circledast} \neq E_A$, $\mathrm{U}(E_A, x_B^{\circledast}, 1) > \mathrm{U}(E_A, E_A, 0)$. By continuity and monotonicity of $\mathrm{U}(x_A, x_B, \lambda)$ in x_A, there exists $p' > 0$ such that $\mathrm{U}(E_A, x_B^{\circledast}, 1) > \mathrm{U}(E_A - p', x_B^{\circledast}, 1) > \mathrm{U}(E_A, E_A, 0)$. Thus, there exists $p' > 0$ such that A pays p'. By Lemma 1, since A pays a positive price at least once, her willingness to pay $\bar{p}_{\mathrm{I}}$ is strictly positive.

Now we show the opposite direction: if $\bar{p}_{\mathrm{I}} > 0$, then A does not choose E_A for B at $p = 0$. Since $\bar{p}_{\mathrm{I}} > 0$, by continuity and monotonicity of $\mathrm{U}(x_A, x_B, \lambda)$ in x_A, there exists some positive price $p < \bar{p}_{\mathrm{I}}$, such that $\mathrm{U}(E_A - p, x_B^*, 1) > \mathrm{U}(E_A, E_A, 0)$, where $x_B^* = \underset{x_B \in [0,E_B]}{\operatorname{argmax}} \mathrm{U}(E_A - p, x_B, 1)$ and $x_B^* \neq E_A$. By monotonicity, we have that $\mathrm{U}(E_A, x_B^*, 1) > \mathrm{U}(E_A - p, x_B^*, 1) > \mathrm{U}(E_A, E_A, 0)$. In other words, at $p = 0$, player A is better off when she chooses x_B^* as player

[32]For example, competitive or spiteful preferences and social-welfare preferences as in Levine (1998), Charness and Rabin (2002) and Cox, Friedman and Gjerstad (2007) are in this category.

[33]Preferences for equality in Fehr and Schmidt (1999) and Bolton and Ockenfels (2000) are example that fall into this category of preferences.

B's payoff than when she chooses E_A. Thus, at a price of 0, player A pays and does not choose E_A for player B.

The second statement of the proposition follows directly from the fact that $\mathrm{U}(x_A, x_B, \lambda) = \mathrm{U}(x_A, x_B, \lambda')$. Thus, for all x_A, x_B, $\mathrm{U}(x_A, x_B, 1) = \mathrm{U}(x_A, x_B, 0)$. Therefore, the Part I logic holds for Part II and $\bar{p}_{\mathrm{II}} = \bar{p}_{\mathrm{I}}$. □

Proposition 3 shows that players whose preferences fit within a large class of social preferences, e.g., social-welfare preferences or competitive preferences, have a strictly positive willingness to pay in Part I of the Power Game. Their willingness to pay, $\bar{p}_{\mathrm{I}}$, depends on the strength of their preferences with respect to the payoff of player B as well as on the available payoff options for B, i.e. the interval $[0, E_B]$. Moreover, in Part II such players choose to implement the same allocations as in Part I and thus pay the same positive prices, i.e. $\bar{p}_{\mathrm{I}} = \bar{p}_{\mathrm{II}}$.

Corollary 1. *For a player with social preferences and no power preferences, $\bar{p}_I = 0$ if and only if such a player chooses E_A for player B when $p = 0$. Moreover, $\bar{p}_{II} = \bar{p}_I = 0$.*

Corollary 1 is the transpose of Proposition 3 and so is directly implied by it. One direct implication of Corollary 1 is that players with preferences for equality should never pay positive prices in either Part of the Power Game.

Finally, we consider subjects who have both social and power preferences. We add a last assumption that states that preferences for power enter the utility function in an additively separable manner. While this assumption is convenient, it is not strictly necessary: all that is required is that x_B^* be independent of λ.[34]

Assumption 5. $\mathrm{U}(x_A, x_B, \lambda) = \mathrm{V}(x_A, x_B) + f(\lambda)$.

Proposition 4. *For a player with power and social preferences, $\bar{p}_I > 0$ and $0 \leq \bar{p}_{II} < \bar{p}_I$.*

Proof. Let's first consider the social component of A's utility function. By Assumptions 4 and 5, for a given payoff of player A, there is a unique optimal allocation for B that is independent of her preferences for power. That is, for every λ, λ', $x_B^* = \underset{x_B \in [0, E_B]}{\mathrm{argmax}} \mathrm{U}(x_A, x_B, \lambda) = \underset{x_B \in [0, E_B]}{\mathrm{argmax}} \mathrm{U}(x_A, x_B, \lambda') = \underset{x_B \in [0, E_B]}{\mathrm{argmax}} \mathrm{V}(x_A, x_B)$.

In particular, if her payoff is $E_A - p$, the allocation for B that maximizes her utility is the same in Part I and in Part II. By Proposition 3 and Corollary 1, a player with a social component to her utility is willing to trade-off $p_s \geq 0$ to maximize her utility and obtain the $(E_A - p_s, x_B^*)$ allocation instead of the (E_A, E_A) allocation, i.e. $\mathrm{V}(E_A - p_s, x_B^*) \geq \mathrm{V}(E_A, E_A)$, where $x_B^* = \underset{x_B \in [0, E_B]}{\mathrm{argmax}} \mathrm{V}(E_A - p_s, x_B)$.

In a setting of increased power, her choice for B is identical and her utility is then $\mathrm{V}(E_A - p_s, x_B^*) + f(1)$ which is strictly greater than $\mathrm{V}(E_A - p_s, x_B^*) + f(0)$, because she enjoys

[34] We choose to work with separability to keep in line with the literatures on outcome- and intentions-based social preferences. See for example Fehr and Schmidt (1999), Charness and Rabin (2002) and Levine (1998), Cox, Friedman and Gjerstad (2007).

power and thus by Assumption 2, $f(\lambda)$ is increasing in λ. By continuity and monotonicity of $\mathrm{V}(x_A, x_B)$ in x_A, there exists $p' > p_s$ such that $\mathrm{V}(E_A - p_s, x_B^*) + f(1) > \mathrm{V}(E_A - p', x_B^*) + f(1) > \mathrm{V}(E_A - p_s, x_B^*) + f(0) \geq \mathrm{V}(E_A, E_A) + f(0)$. Equivalently, $\mathrm{U}(E_A - p_s, x_B^*, 1) > \mathrm{U}(E_A - p', x_B^*, 1) > \mathrm{U}(E_A - p_s, x_B^*, 0) \geq \mathrm{U}(E_A, E_A, 0)$. Thus, A is willing to pay more to implement both her social and power preferences than to implement her social preferences only. As a result A is willing to pay more in Part I than in Part II.[35] $\square$

Proposition 4 states that if a player enjoys power and has social preferences, she is willing to sacrifice a larger fraction of her payoff in order to both obtain power and implement her social preferences as in Part I than in order to only implement her social preferences, as in Part II.

[35]If a player is averse to power, then the inequalities are reversed and $\bar{p}_{\mathrm{II}} > \bar{p}_{\mathrm{I}}$, and $\bar{p}$ may be negative.

APPENDIX E.

In this Appendix we show that when we use the whole sample and allow any number of skips, we obtain the same results. For that, we redo all Figures and Tables that appear in Sections 3.2, 3.3, and 3.4 of the Main text, but using all 288 subjects. 178 subjects make no skips across Parts I and II of the Power Game, 37 subjects skip one price, 16 subjects make two skips, 17 subjects make three skips, and 40 subjects make four skips or more. For those subjects who make no skips we calculate their willingness to pay in the same way as in the main text. For those subjects who make skips, a subject's willingness to pay is defined as the maximum price p at which she pays before making her first skip, or -\$0.25 if she does not pay at a price of -\$0.25.[36]

Figure D1 shows the joint distribution of the willingnesses to pay in Parts I and II, $\bar{p}_{\mathrm{I}}$ and $\bar{p}_{\mathrm{II}}$. It complements Figure 6 from the main text. Figure D2 shows the proportion of subjects with various preference classes based on their willingnesses to pay in Parts I and II, $\bar{p}_{\mathrm{I}}$ and $\bar{p}_{\mathrm{II}}$ and it complements Figure 7 of the main text. Note that in the entire sample we obtain a similar fraction of Power+ subjects (26%) as we did in the well-behaved sample (27.5%).

Table D1 presents evidence that our preference classification is not affected by subjects' confusion or mistakes. In the final questionnaire, we asked subjects whether they found anything in the experiment confusing (see Appendix C). The second column of Table D1 we presents the fraction of subjects whose answer was positive, regardless of what they indicated was confusing. Note that Power+ subjects are not more confused about the game than others.

To demonstrate that our preference classification is not driven by skips, we calculate the total number of skips in Part I, II, and I* by preference class and report the results in Table D1.[37] Using a series of pairwise Kolmogorov-Smirnov and Wilcoxon-Mann-Whitney tests, we can reject the argument that Power+ subjects are simply those who make more mistakes. In fact, they make fewer mistakes than the subjects with social preferences (all together), and are no different than those subjects in the Standard class. For the former comparison, the p-values in a Kolmogorov-Smirnov and in a Wilcoxon-Mann-Whitney test are 0.01 and 0.001 respectively. For the latter comparison, both p-values are greater than 10%. Thus, it is not the case that subjects in the Power+ preference class are simply those who make more skips than others.

Figure D3 shows the cumulative distribution function of the amounts given to B in Part I, averaged per subject, separately for subjects in the Power+ preference class and those with social preferences (Social, Social&Power+, and Social&Power-). This figure complements Figure 8 of the main text. Figure D4 presents the distribution of preference classes based on subjects' paying behavior in Parts I* and II and it complements Figure 9 of the main text. Finally, Figure D5 presents the fraction of subjects in each preference class, as obtained from Parts I and II, who retain that preference class defined using Parts I* and II. This figure complements Figure

[36] We can also define subject's willingness to pay as the maximum price she ever pays. Our results are unchanged if we do so. These results are available upon request.

[37] Our results are unchanged if we compare the total number of skips in Parts I and II across different preference classes.

10 of the main text. In Table D2 we compare our subjects to those of Charness and Rabin (2002) and Chen and Li (2009) by presenting the proportion of subjects who choose the first option in decision problems CR1-CR6. This table complements Table 4 of the main text. In Table D3 we compare subjects' behavior in the Charness and Rabin (2002) task across subjects with different preferences: Standard, Power+, and social preferences (all together). This table complements Table 5 of the main text. We run a series of tests of proportions comparing the fraction of subjects who choose the first payoff-maximizing option for the Standard and Power+ subjects as well as for the Power+ and Social (all together) subjects. The last two columns of Table D3 report the corresponding p-values. Just as in the main text, our results demonstrate that Power+ subjects are no different than Standard ones and choose the payoff-maximizing option significantly more often than subjects with social preferences.

Aggregating behavior for each subject across decision problems CR2-CR5, and PT1-PT2, we find that the fraction of subjects who always choose the payoff-maximizing option among those with standard preferences is 46.8%. For Power+ subjects this fraction is 48.0% , statistically indistinguishable from that one for the Standard preference class. Finally, for the subjects with social preferences this fraction is 9.0%, and it is significantly lower than those fractions for the subjects with either Power+ or Standard preferences.

Finally, Table D4 compares subjects' paying and giving behavior between Parts I and I* for those subjects with stable preferences. It complements Table 6 of the main text.

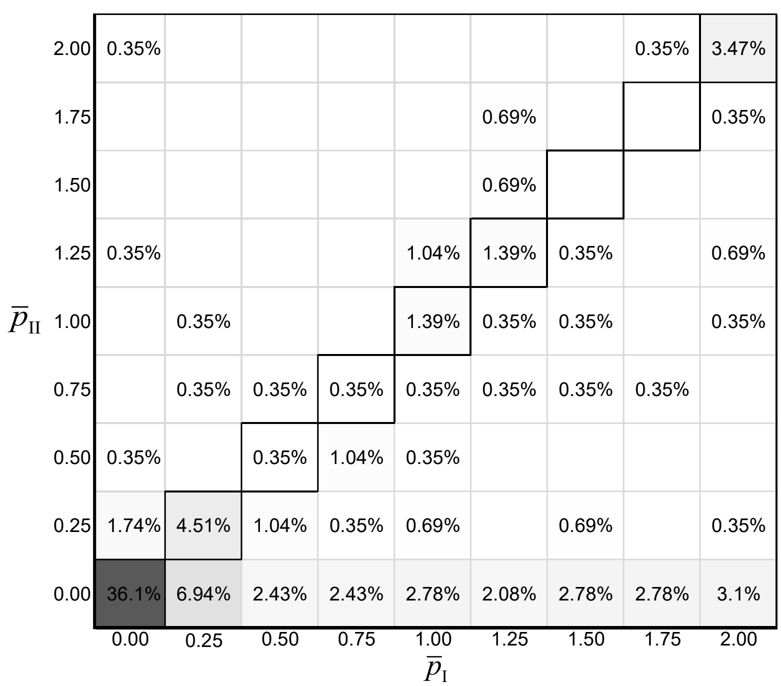

FIGURE D1. Joint distribution of the willingnesses to pay in Parts I and II, $\bar{p}_{\mathrm{I}}$ and $\bar{p}_{\mathrm{II}}$.

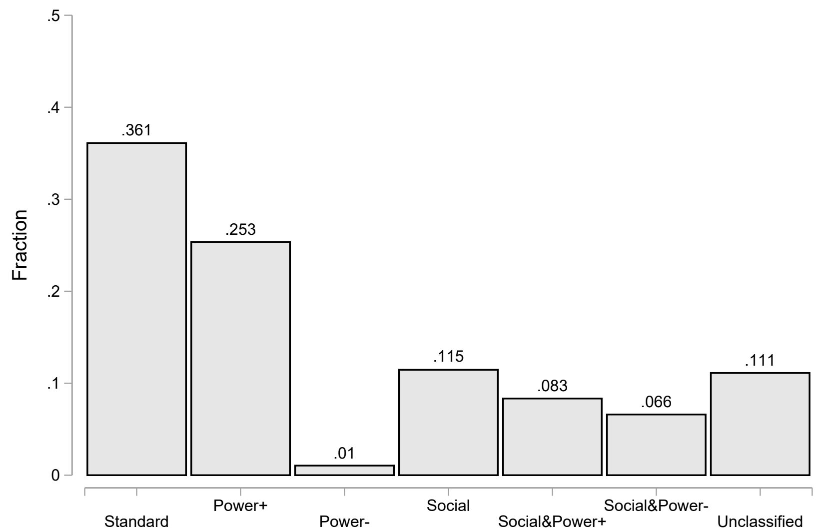

FIGURE D2. Distribution of preference classes based on the willingnesses to pay in Parts I and II, $\bar{p}_{\mathrm{I}}$ and $\bar{p}_{\mathrm{II}}$.

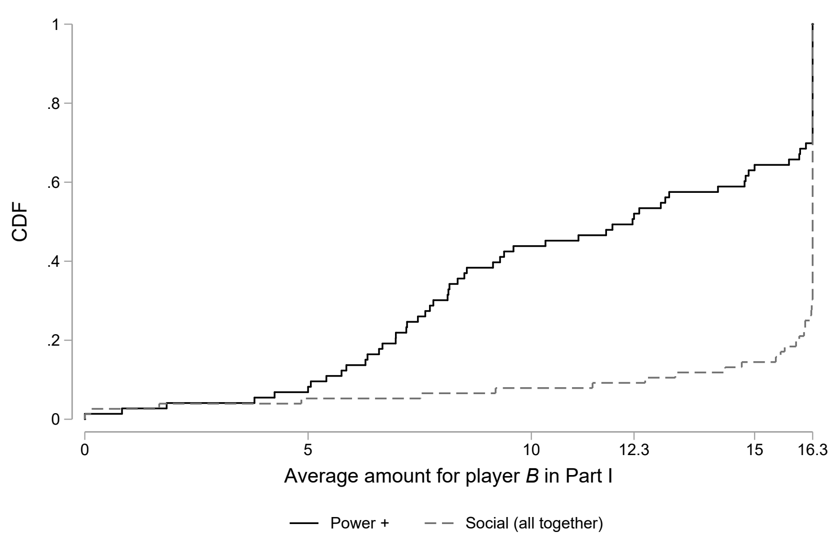

FIGURE D3. Distribution of the average amount given to player B in Part I

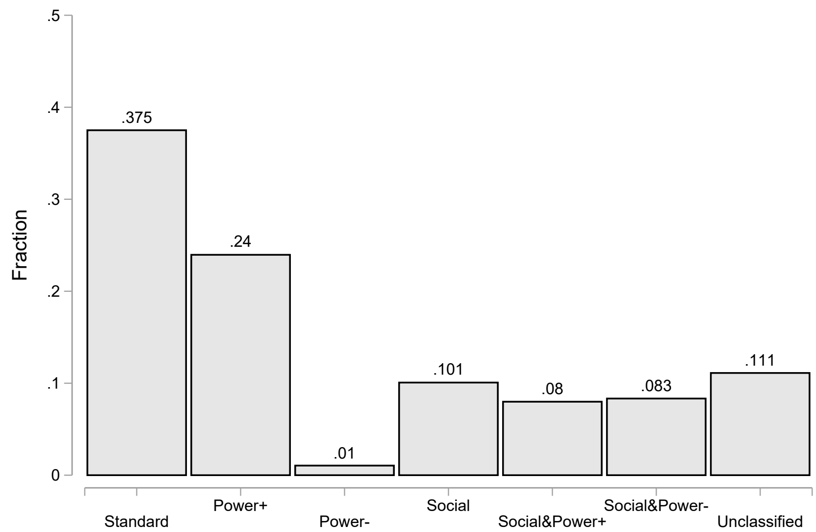

FIGURE D4. Distribution of preference classes based on subjects' paying behavior in Parts I* and II.

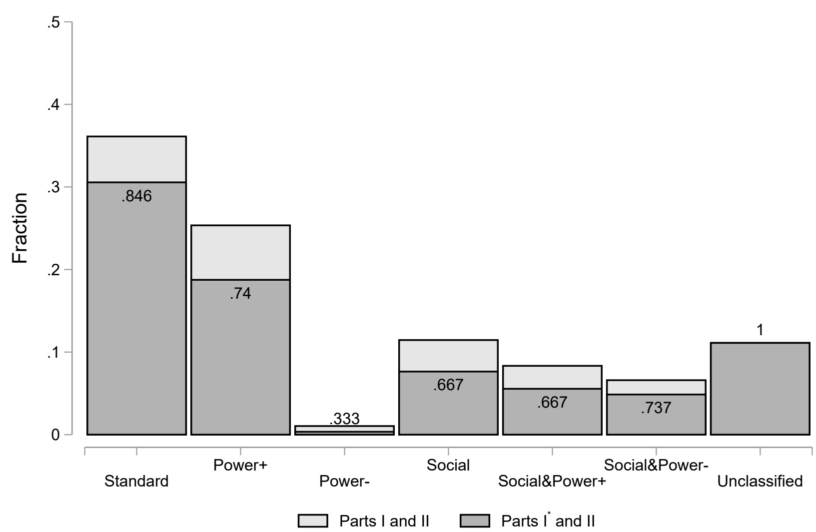

FIGURE D5. Fraction of subjects in each preference class, as obtained from Parts I and II, who retain their preference class, as obtained using Parts I* and II.

TABLE D1. Confusion and the total number of skips across Parts I, II and I* of the Power Game by preference class for the entire sample of 288 subjects.

		Confusion[a]	Total number of skips in Parts I, II and I*						
Preferences Class	Subjects	Mean	Mean	St.Dev.	Min.	p25	Median	p75	Max.
Standard	104	0.14	0.83	2.15	0	0	0	0	12
Power +	73	0.15	1.07	2.14	0	0	0	1	13
Power −	3	0.00	6.33	5.51	0	0	9	10	10
Social (All)	76	0.14	1.67	2.00	0	0	1	3	8
Social	33	0.18	1.45	2.24	0	0	0	2	8
Social & Power +	24	0.13	1.63	1.58	0	0	1	3	5
Social & Power −	19	0.11	2.11	2.05	0	0	2	3	6
Unclassified	32	0.44	4.13	4.40	0	0.5	3	6	17

[a]Confusion is a simple dummy that equals one if in the final questionnaire a subject indicated that he/she found anything confusing about the study.

TABLE D2. Fraction of subjects choosing the first option in the Charness and Rabin (2002) task across three samples: Charness and Rabin (2002), Chen and Li (2009), and our sample.

Decision	First Option	Second Option	CR2002	CL2009	Our Subjects[a]
CR1	(6.60, 6.60)	(6.60, 12.30)	31%	33%	24%
CR2	(6.60, 6.60)	(6.20, 12.30)	51%	82%	50%
CR3	(10.50, 5.30)	(8.80, 12.30)	67%	76%	77%
CR4	(12.30, 3.50)	(10.50, 10.50)	27%	50%	57%
CR5	(12.30, 0.00)	(6.15, 6.15)	78%	64%	74%
CR6	(3.10, 12.30)	(0.00, 0.00)	100%	NA	99%

[a]The sample includes 288 subjects who make any number of skips across Parts I, II, and I* of the Power Game.

TABLE D3. Fraction of subjects choosing the first option in the independent decision problems in Part II by preference class.

	Option		Preference Class			p-value	
Decision	First	Second	Standard	Power +	Social (All)	P+ vs. Std	P+ vs. Soc
CR1	(6.60, 6.60)	(6.60, 12.30)	20%	38%	13%	0.0140	0.0004
CR2	(6.60, 6.60)	(6.20, 12.30)	63%	65%	18%	0.7517	0.0000
CR3	(10.50, 5.30)	(8.80, 12.30)	88%	84%	60%	0.4003	0.0003
CR4	(12.30, 3.50)	(10.50, 10.50)	69%	71%	30%	0.8499	0.0000
CR5	(12.30, 0.00)	(6.15, 6.15)	81%	79%	63%	0.6857	0.0141
PT1	(10.10, 5.20)	(9.10, 9.10)	70%	75%	36%	0.5122	0.0000
PT2	(12.30, 5.10)	(10.10, 12.30)	83%	84%	45%	0.8411	0.0000
Sanity checks							
CR6	(3.10, 12.30)	(0.00, 0.00)	99%	99%	98%	0.7791	0.6635
PT3	(12.55, 12.80)	(12.30, 12.30)	94%	93%	90%	0.9218	0.4315
PT4	(12.30, 9.60)	(9.60, 12.30)	99%	97%	91%	0.3484	0.0919
PT5	(12.30, 7.80)	(7.80, 5.40)	100%	99%	98%	0.2225	0.6635
PT6	(6.15, 6.15)	(0.00, 0.00)	100%	100%	99%	1.0000	0.3572

TABLE D4. Subjects' behavior in Parts I and I*, for those subjects who retain their preference class in Part I*.

Panel A: Subjects' paying behavior

Preference Class	Subjects	Max Paid Part I	Max Paid Part I*	Paid the Same in Parts I and I*
Standard	93	0.00	0.00	1.00
Power +	56	1.15	1.01	0.62
Social (All)	60	0.97	0.97	0.64
Social	23	1.09	1.09	1.00
Social & Power +	19	1.25	1.07	0.32
Social & Power −	18	0.53	0.67	0.53

Panel B: Subjects' giving behavior

Preference Class	Subjects	Payoff for *B* Part I	Payoff for *B* Part I*	Gave the Same in Parts I and I*
Standard	93	13.87	14.18	0.66
Power +	56	10.72	11.11	0.23
Social (All)	60	14.59	14.96	0.76
Social	23	15.82	15.76	0.96
Social & Power +	19	14.35	14.39	0.68
Social & Power −	18	13.27	14.49	0.59

APPENDIX F.

Below we present the instructions that the subjects received in our vignette study.

PREAMBLE

You will read descriptions of five different situations where one individual has power over another individual or group of individuals. For each of these scenarios, you will be asked to answer one question about the amount of power the individual in charge has over others. There are no right or wrong answers for these questions and we simply ask you to answer the questions based on your feelings about the situation and to the best of your judgment.

STUDY 1

Harbor Inc. is a new furniture company that manufactures and delivers custom made sofas that individuals can view and order online. It is currently developing a website that it expects to be instrumental in facilitating online orders and ensuring sales growth. All the employees are encouraged to give their feedback on the website and are offered spot rewards for useful comments and suggestions. The company CEO Andre Redway is considering what would be the best way to reward the employees for their feedback. One option is to give a bonus of $50 to all employees who submit a proposal independent of the proposal quality. Another option is to award a bonus between $15 and $75 based on the proposal quality. After some deliberation, Andre announces that he'll award a fixed bonus of $50 to everyone who submits a feedback proposal. [Condition 2: After some deliberation, Andre announces that he'll award bonuses of variable value between $15 and $75 based on proposal quality to everyone who submits a feedback proposal.]

On a scale from 1 to 5, indicate how much power Andre has over the bonuses of his employees *after* the announcement has been made.

1 No power at all
2 Low amount of power
3 Intermediate amount of power
4 High amount of power
5 Very high amount of power

STUDY 2

Prof. Antony Croxton is designing a new course in food microbiology for graduate students. The course will cover the role of microorganisms in food infection and spoilage as well as basic principles in contamination control and germicidal treatment during food processing and distributing for consumption. The course will be offered as an elective for 2nd year master students and it is expected that about 10-12 students will sign up for the course. Usually such advanced courses are graded on a pass/fail basis. However, given the small number of students prof. Croxton is considering whether to grade the course using numerical grades, from 0 to 100, and to provide the students with a better final assessment of their knowledge. In the end, prof. Croxton decides that the course will be graded on a pass/fail basis. [Condition 2: At the end, prof. Croxton decides that the course will be graded on a pass/fail basis.]

On a scale from 1 to 5, indicate how much power prof. Croxton has over the grades of the students *after* the grade scheme decision has been made.

1 No power at all
2 Low amount of power
3 Intermediate amount of power
4 High amount of power
5 Very high amount of power

STUDY 3

Carla Thompson works as a logistics manager for Pecker Outdoors, a footwear company. She oversees the storage and distribution of finished products and ensures that customers receive products on time. Her team of 11 people includes warehouse and transportation workers, inventory control employees and customer service employees. Her responsibilities include providing coaching and performance feedback to all team members, including new hires, monitoring individual and team metrics and performance versus targets, ensuring quality standards for all processes as well as evaluating overall team and individual performance of her subordinates and delivering those performance reviews to upper management. As part of her annual performance review, Carla usually makes a recommendation on cash bonuses to be paid to her team members. The bonus amount can vary between 0 and 20 percent of the employee annual salary. Historically, Carla always recommended the same bonus of 10 percent to everyone on her team. This year, despite having a few slackers and a few star employees on the team, Carla's recommendation was to pay the same bonuses to all of her team members. [Condition 2: This year, due to having a few slackers and a few star employees on the team, Carla's recommendation was to pay higher bonuses to better performing employees and lower bonuses to the slackers.]

On a scale from 1 to 5, indicate how much power Carla had over the compensation of her subordinates.

1 No power at all
2 Low amount of power
3 Intermediate amount of power
4 High amount of power
5 Very high amount of power

Study 4

Angela Buzzello is a teaching assistant for Prof. John Kasiz. Prof. Kasiz teaches an introductory class in accounting for undergraduate students. He asks Angela to give tutorials for the course, where she answers student questions and solves problems provided by the professor. Angela does very well in the tutorials and prof. Kasiz is thinking whether to ask her to grade the final exam. Usually it is a professor responsibility to grade final exams while teaching assistants grade all other assignments. However, Angela is so comfortable with the course material and by now knows all the students personally, that prof. Kasiz has decided to delegate all the grading to her. He does tell the students that their final exam will be graded by Angela. [Condition 2: However, he does not tell the students that their final exam will be graded by Angela.]

On a scale from 1 to 5, indicate how much power Angela has over the grades of the students.

1 No power at all
2 Low amount of power
3 Intermediate amount of power
4 High amount of power
5 Very high amount of power

STUDY 5

George Hansen is an inventory control manager and he is looking to fill in a job opening for a Team Lead. Team lead will be responsible for keeping detailed records of inventory use and sales, and advising George on ordering where necessary. There are five suitable candidates among the inventory control employees and George considers each of them carefully. Historically, such within-team promotions were decided by an immediate manager such as George and then simply rubber stamped by the upper management. However, given that the company recently experienced several major upgrades in its operation and distribution processes, George's boss Eliza decides to review the credentials of the candidates herself. She asks George to write a detailed memo that includes descriptions of strong and weak sides for each candidates' and George's justification for his final recommendation on this promotion. Eliza is considering whether to inform the workers or not about such an intervention because she doesn't want to undermine George's immediate authority. At the end, she decides to inform the workers. [Condition 2: At the end, she decides not to inform the workers.]

On a scale from 1 to 5, indicate how much power George had over the promotion decision.

1 No power at all
2 Low amount of power
3 Intermediate amount of power
4 High amount of power
5 Very high amount of power

APPENDIX G.

In this Appendix we present the results we obtained using our previous experimental design. We had 16 sessions with a total of 292 subjects. Among those, 128 or 43.8% were well-behaved in the sense that they made no skips in Parts I and II of the Power Game (we did not run Part I* in those sessions). Using those well-behaved subjects, we replicate Figures 7 and 8 and Tables 4 and 5 from the Main text.

In the analysis below we emphasize the following:

(1) Our previous design did not allow us to distinguish between subjects with Power- and Standard preference classes. The results presented in the Main text suggest Power- subjects are rare (only one subject out of 163 exhibited behavior consistent with such preferences). Thus, we assign subjects who never paid in our previous experiment, even at a price of zero, to the Standard preference class.

(2) As in the Main text, we use all well-behaved subjects, with no exception.

In Figure F1 we show the distribution of preference classes based on subjects' willingnesses to pay in Parts I and II. This figure complements Figure 7 from the Main text. Just as is the case with the current design, Power+ subjects represent a substantial fraction of the population. In fact, under this older design, they are the most common class.

In Figure F2 we show the distribution of the average amount given to player B by preference class. This figure complements Figure 8 from the Main text. Just as is the case with the current design, Power+ subjects give to B in a very variable way, while those identified as having social preferences show almost no variability, either across or within subjects. Kolmogorov-Smirnov and Wilcoxon-Mann-Whitney tests show that the distribution of amounts given by subjects in the Power+ preference class is statistically different than that of those who have social preferences, with p-values less than 0.001. The unit of observation is the average amount given by each subject. Moreover, the within-subject mean standard deviation of choices for B's payoff is higher for subjects in the Power+ than in the Social (all together) preference class: 1.72 versus 0.15. A two-sided test comparing these standard deviations shows that they are statistically different with a p-value less than 0.001.

In Table F1 we compare our subjects to those of Charness and Rabin (2002) and Chen and Li (2009) by presenting the proportion of subjects who choose the first option in decision problems CR1-CR6. This table complements Table 4 from the Main text and presents results very similar to those in the Main text.

Finally, in Table F2 we compare our subjects' choices to those of Charness and Rabin (2002) by presenting the proportion of subjects who choose the first option in decision problems CR1-CR6, PT1-PT2, and PT4-PT6 (decision problem PT3 was not included in our previous design). We run a series of tests of proportions comparing the fraction of subject choosing the first option for the Standard and Power+ subjects as well as for the Power+ and Social (all together) subjects. As in Table 5 of the Main text, we report the p-values associated with those tests in the last two columns of Table F2. Our results are robust across the two experimental designs. For example, when we aggregate behavior for each subject across decision problems CR2-CR5, and PT1-PT2, we find that the fraction of subjects who always choose the payoff-maximizing option among those with standard preferences is 56.8%. For Power+ subjects this fraction is 46.3%, statistically indistinguishable from those in the Standard preference class, and it is significantly higher than 6.9%, the fraction for those who have social preferences. The p-value for a two-sided test of proportions comparing the first two is 0.300, and the p-values for each pairwise comparison between social and the others are less than 0.001.

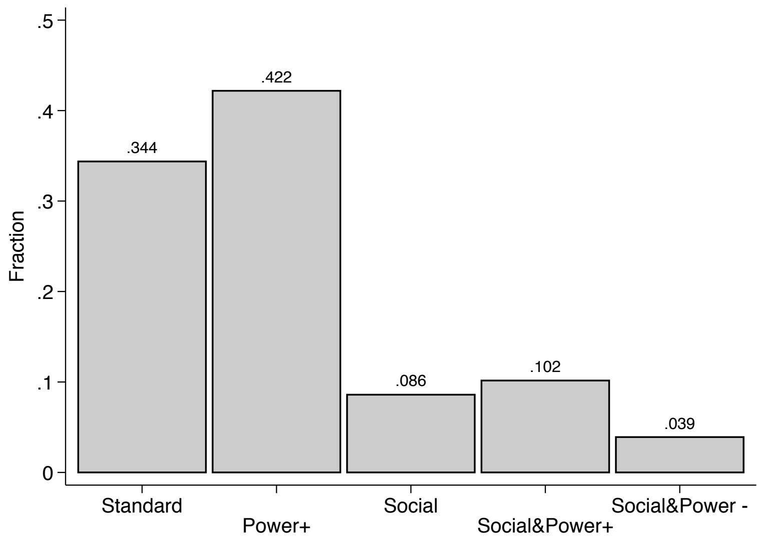

FIGURE F1. Distribution of preference classes, obtained using our previous design.

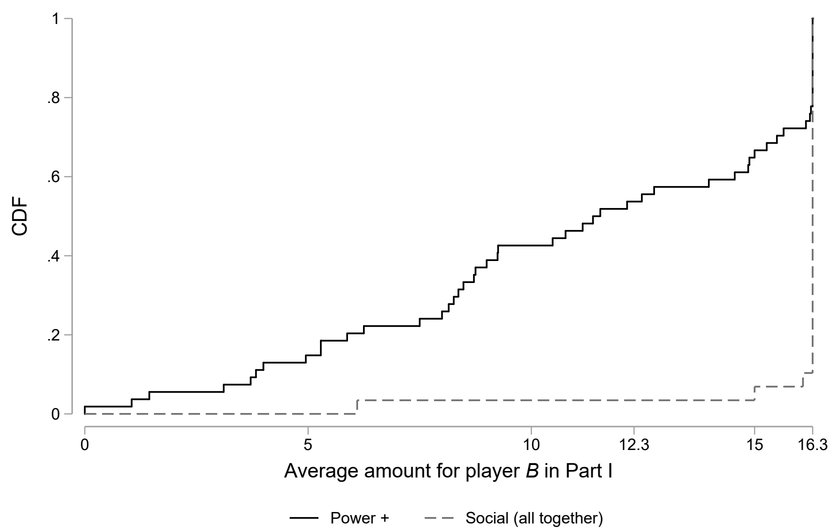

FIGURE F2. Distributions of the average amount given to player B by preference class, obtained using our previous design.

TABLE F1. Fraction of subjects choosing the first option in the Charness and Rabin (2002) task across three samples: Charness and Rabin (2002), Chen and Li (2009), and our sample of 128 well-behaved subjects.

Decision	First Option	Second Option	CR2002	CL2009	Our Subjects
CR1	(6.60, 6.60)	(6.60, 12.30)	31%	33%	32%
CR2	(6.60, 6.60)	(6.20, 12.30)	51%	82%	56%
CR3	(10.50, 5.30)	(8.80, 12.30)	67%	76%	84%
CR4	(12.30, 3.50)	(10.50, 10.50)	27%	50%	66%
CR5	(12.30, 0.00)	(6.15, 6.15)	78%	64%	80%
CR6	(3.10, 12.30)	(0.00, 0.00)	100%	NA	97%

TABLE F2. Fraction of subjects choosing the first option in the independent decision problems in Part II by preference class, obtained using our previous design.

	Option		Preference Class			p-value	
Decision	First	Second	Standard	Power +	Social (All)	P+ vs. Std	P+ vs. Soc
CR1	(6.60, 6.60)	(6.60, 12.30)	27%	50%	7%	0.0222	0.0000
CR2	(6.60, 6.60)	(6.20, 12.30)	68%	74%	7%	0.5207	0.0000
CR3	(10.50, 5.30)	(8.80, 12.30)	87%	85%	76%	0.6164	0.1463
CR4	(12.30, 3.50)	(10.50, 10.50)	78%	83%	28%	0.4414	0.0050
CR5	(12.30, 0.00)	(6.15, 6.15)	84%	87%	72%	0.4317	0.5642
PT1	(10.10, 5.20)	(9.10, 9.10)	77%	69%	52%	0.3346	0.0658
PT2	(12.30, 5.10)	(10.10, 12.30)	88%	91%	28%	0.5705	0.0793
Sanity checks							
CR6	(3.10, 12.30)	(0.00, 0.00)	95%	96%	100%	0.8341	0.2941
PT4[a]	(12.30, 9.60)	(9.60, 12.30)	95%	98%	100%	0.4413	0.4610
PT5	(12.30, 7.80)	(7.80, 5.40)	98%	94%	100%	0.4140	0.1961
PT6	(6.15, 6.15)	(0.00, 0.00)	100%	100%	100%	1.0000	1.0000

[a]Decision problem PT3 was not included in our previous design.