

Conveniently Upset: Avoiding Altruism by Distorting Beliefs about Others' Altruism[†]

By RAFAEL DI TELLA, RICARDO PEREZ-TRUGLIA, ANDRES BABINO,
AND MARIANO SIGMAN*

We present results from a “corruption game” (a dictator game modified so that recipients can take a side payment in exchange for accepting a reduction in the overall size of the pie). Dictators (silently) treated to be able to take more of the recipient’s tokens, took more of them. They were also more likely to believe that recipients had accepted side payments, even if there was a prize for accuracy. The results favor the hypothesis that people avoid altruistic actions by distorting beliefs about others’ altruism. (JEL C72, D63, D64, D83)

He who wants to kill his dog, accuses it of rabies.
Molière¹

Sometimes, the actions we enjoy taking have a negative effect on other people. Because holding these people in high opinion reduces the pleasure derived from such actions, it is useful to change these opinions. Consider, for example, the case of the president of a powerful country that wishes to take control of a weaker country’s natural resources. To justify an invasion, he comes to believe the leader of the weaker country is in contact with a hostile terrorist network and has developed weapons of mass destruction.² Or consider the case of a man who would like to cheat on his wife and goes on to believe that she often mistreats him. Holding more

*Di Tella: Harvard Business School, 15 Harvard Way, Morgan Hall 283, Boston, MA 02163 (e-mail: rditella@hbs.edu); Perez-Truglia: Microsoft Research, New England Research and Development (NERD) Lab, 1 Memorial Drive, Office 12073, Cambridge, MA 02142 (e-mail: rtruglia@microsoft.com); Babino: Departamento de Física, UBA, Caldas 1700 3, Ciudad Autónoma de Buenos Aires, C.P. 1426, Argentina (e-mail: ababino@df.uba.ar); Sigman: Departamento de Física, FCEN, UBA, and IFIBA, and Universidad Torcuato Di Tella, Almirante Juan Saenz Valiente 1010, C1428BII, Buenos Aires, Argentina (e-mail: msigman@utdt.edu). We thank Fiorella Benedetti, Tamara Niella, and Micaela Sviatschi for excellent research assistance and Nageeb Ali, Roland Bénabou, Nyla Branscombe, Alex Haslam, James Konow, Julio Rotemberg, and Eldar Shafir for many helpful comments. We also thank very useful feedback and suggestions from three anonymous referees. Rafael Di Tella thanks the support of the Canadian Institute for Advanced Research, and Mariano Sigman thanks CONICET and the James McDonnell Foundation 21st Century Science Initiative in Understanding Human Cognition. The authors declare that they have no relevant or material financial interests that relate to the research described in this paper. This is a revised version of Di Tella and Perez-Truglia (2010).

[†]Go to <http://dx.doi.org/10.1257/aer.20141409> to visit the article page for additional materials and author disclosure statement(s).

¹From *Les Femmes savantes* (1672), “Qui veut noyer son chien l’accuse de la rage” (translated by the authors).

²Some observers claim that the US-led invasion of Iraq in 2003, and the following opening up of the country’s oil industry to western companies, fits this description. Apparently, General John Abizaid, former head of US Central Command and Military Operations in Iraq in 2007, explained, “Of course it’s about oil; we can’t really deny that.” (see Antonia Juhasz, “Why the War in Iraq Was Fought for Big Oil,” CNN, April 15, 2013.)

negative beliefs about his wife may not be easy, but doing so is convenient because it reduces any guilt he may feel. Finally, consider the case of an employee who can exert low effort at work at the cost of letting down a boss who is nice and fair. Tired of this constraint inducing her to exert costly effort, the employee can convince herself that the boss is not that nice after all.

A key difficulty in studying self-serving biases empirically is reverse causality: people are likely to behave selfishly toward people whom they already hold in low esteem. In terms of the marital example above, a man who believes his wife mistreats him is more likely to go on to have an affair. Our approach involves designing a series of laboratory experiments where individuals face different incentives to act selfishly and then measuring how this affects their beliefs about others and their actions. Some of these experiments also allow us to study how ambiguity regarding the recipient's motivations promotes selfish actions.

We study beliefs and choices in a modified dictator game. In the standard dictator game, one subject, called the dictator, decides how to divide a fixed sum of money with an anonymous counterpart, called the recipient (e.g., Forsythe et al. 1994). We introduce a variation on this design, which we call the "corruption game" and which consists of two stages. In the first stage, all subjects complete a set of tasks that generate tokens, which can be converted into money. In the second stage, subjects make their choices: the dictator individually decides how to allocate the tokens between himself and the recipient and the recipient chooses the price at which both subjects can cash in the tokens. The recipient chooses either a high price (e.g., \$2 per token) or a low price (e.g., \$1 per token), whereby choosing a low price additionally gives the recipient a side payment (e.g., \$10).³ Moves are simultaneous: the dictator (whom we call the "allocator") allocates tokens without knowing the price chosen by the recipient, and the recipient (whom we call the "seller") chooses a price without knowing the allocator's decision.

The first key manipulation of our design is that we vary the allocator's ability to be selfish. We do so by randomizing the maximum number of tokens that an allocator can appropriate. In the $Able=8$ treatment, the allocator is able to take up to 8 of the seller's tokens for himself (or give up to 8 of her own tokens to the seller), i.e., each subject is guaranteed at least 2 tokens. In the $Able=2$ treatment, the allocator can transfer at most 2 tokens to or from the seller, i.e., each is guaranteed at least 8 tokens. The outcome of this randomization is only observable to the allocator. Given that the sellers cannot observe the treatment (i.e., whether the allocator can take 2 or 8 tokens), all allocators should expect the same behavior from sellers. However, if self-deception were possible, allocators who can take more tokens from the seller (i.e., $Able=8$ instead of $Able=2$) have more incentives to convince themselves that the seller is unkind (i.e., accepts a low price for the tokens in exchange for a side payment). To test this hypothesis, we ask allocators to guess the proportion of the population of sellers that accepts side payments, with an opportunity to get a reward if they guess correctly. As hypothesized, we find that allocators who

³ Some corruption episodes have a similar payout structure (e.g., when an agent under-invoices a sale and keeps the difference). There is a small experimental literature on corruption that studies other aspects of the problem. A prominent example is the work of Abbink, Irlenbusch, and Renner (2002), using variations of the trust game (see Berg, Dickhaut, and McCabe 1995). See Dusek, Ortmann, and Lízal (2004) for a review.

have the opportunity to take more tokens convince themselves that the sellers are unkind, reporting that a higher proportion of sellers accept the side payment and sell for a low price.

We conduct three additional games, changing different aspects of the original experiment on a different subject pool. The first new game introduces changes in the design aimed at studying the robustness of our results. In spite of several modifications, we find very similar results. Second, we conduct a variation of the corruption game where the computer chooses the price on the sellers' behalf, eliminating the ambiguity regarding the actions of the seller and, hence, the ability to engage in self-deception. As expected, we find that allocators take fewer tokens when the ambiguity about the seller's actions is eliminated, suggesting that the ability to engage in self-deception does indeed affect the decision to be selfish. Third, we conduct a falsification test based on a variation of the original game where a computer made the allocations. Intuitively, since the computer is responsible for their choice, allocators should not need to deceive themselves into thinking the sellers are unkind. As expected, we find no self-serving bias in this variation of the game.

Our results appear to be economically significant. Our preferred estimates indicate that the incentives we provide increase the allocator's belief regarding the probability that the seller took the "unkind" action by 20 percentage points, and make the allocator take 2.5 additional tokens out of the 10 tokens in the seller's pile.

The rest of the paper proceeds as follows. The Section I discusses the main hypothesis and relates our paper to the existing literature. Section II presents the experimental design and results of our basic "corruption game." Section III presents the results from three variations of the basic design which allow us to address potential confounding factors and test additional hypotheses. Section IV concludes.

I. Main Theoretical Hypotheses and Background

A. Theoretical Hypotheses

Our main hypothesis is that individuals manage their self-image while trying to earn money. Specifically, we study two hypotheses related to self-deception:

HYPOTHESIS 1: *Beliefs about others are affected by people's own desire to be selfish.*

HYPOTHESIS 2: *Selfish actions depend on people's ability to manipulate their beliefs about others.*

In the context of our "corruption" game, Hypothesis 1 predicts that allocators who take more tokens from the seller will have incentives to convince themselves that the sellers acted unkindly. Self-deception is valuable, so this hypothesis predicts that subjects will be willing to take costly actions (e.g., to pay) to maintain these beliefs. Yet, Hypothesis 1 does not necessarily imply that self-deception affects the decision to be selfish. It is possible that allocators make their choice without engaging in self-deception, and only later, when reflecting on their behavior, change their beliefs.

By contrast, Hypothesis 2 focuses on actions and is, therefore, stronger. It suggests that allocators who are able to engage in self-deception will be more selfish. That is, if they can convince themselves that the sellers are unkind, they can allow themselves to be more selfish and take more tokens, while at the same time maintaining the view that they are fair (see the Appendix for a simple model, and Di Tella and Perez-Truglia 2010 for an alternative approach).

B. Relation to Previous Work

This builds on a large literature which studies fairness in games. For example, a standard interpretation in papers finding significant sharing in the dictator game is that people want to think of themselves as being fair,⁴ or to be perceived as fair by others.⁵ Beliefs play an explicit role in theories of reciprocal fairness, where agents form a belief about other player's altruism so as to respond like with like (see Levine 1998 and Rotemberg 2005, 2008; see also the evidence in Ben-Ner et al. 2004).

The possibility that beliefs exhibit a self-serving bias has been studied since the development of the theory of cognitive dissonance (e.g., Hastorf and Cantril 1954; Festinger 1957).⁶ A classic example is Lerner (1982), which discusses how people tend to believe in a just world, even in the presence of contradictory evidence.⁷ In economics, the possibility of self-serving biases goes back to Adam Smith (Konow 2012). More recently, several studies have demonstrated the presence of self-serving bias and its economic significance. For instance, Babcock, Wang, and Loewenstein (1996) and Babcock and Loewenstein (1997) show that the self-serving bias significantly impacts bargaining behavior, promoting impasse. A striking example is Babcock, Wang, and Loewenstein (1996), which reports that teacher contract negotiators in the United States select "comparable" districts in a biased fashion and that this is correlated with strike activity.⁸

Closer to our work are Rabin (1995) and Konow (2000), which study self-serving biases in the context of fairness concerns.⁹ Rabin (1995) presents a model where an

⁴ See Kahneman, Knetsch, and Thaler (1986); Hoffman et al. (1994); and Bolton, Katok, and Zwick (1998). A vast literature studies different aspects of these preferences, including Rabin (1993); Fehr and Schmidt (1999); Bolton and Ockenfels (2000); Heinrich et al. (2001); and Malmendier, te Velde, and Weber (2014); inter alia.

⁵ See Andreoni and Bernheim (2009). One important related finding is that players' perceived "rights" (to whatever sum is being distributed) heavily influence decisions. In a classic demonstration of this effect, Hoffman and Spitzer (1985) and Hoffman et al. (1994) show that the distribution of payoffs is affected by having players "earn" their roles. See, also, Ruffle (1998); Cherry, Frykblom, and Shogren (2002); and Oxoby and Spraggon (2008). Dal Bó, Foster, and Putterman (2010) find that "democratically" electing the rules of the game affects behavior.

⁶ When two cognitions (e.g., beliefs) are inconsistent, they are said to be "dissonant." See also Akerlof and Dickens (1982) and Oxoby (2003). In the case of dictator games, the dissonant cognitions are the desire to keep the entire pie and to think of oneself as fair. People appear to be motivated to reduce dissonance. Models that explore how individuals selectively recall (or omit) information in a self-serving manner include Rabin and Schrag (1999); Compte and Postlewaite (2004); Köszegi (2006); and Mobius et al. (2014). For a discussion of overconfidence in a Bayesian context, see Benoit and Dubra (2011).

⁷ Bénabou and Tirole (2006) studies how belief-distortion could be a useful motivational strategy, while Caplin and Leahy (2001) and Brunnermeier and Parker (2005) study the consumption value of overoptimistic beliefs.

⁸ In contrast, the survey of tithing practices among Mormons studied in Dahl and Ransom (1999) finds little evidence of the use of self-serving definitions of what constitutes income for charity.

⁹ See also Konow (2003) and Cappelen et al. (2007). The psychology literature on communication has showed how motivated reasoning is constrained by the extent to which reasonable justifications can be invoked (see, for example, Kunda 1990), while work on elastic justification by Hsee (1996) showed that unjustifiable factors influenced decisions more when justifiable factors were more ambiguous. Schweitzer and Hsee (2002) presents evidence suggesting that the reason private information constrains motivated communication is that people will eventually face excessive costs justifying (to themselves) extreme claims about inelastic information. Our results are related to

individual can increase her private consumption but at the cost of harming others. Due to a moral constraint, the individual gets disutility from harming others. Rabin (1995) argues that an individual in such situation could use self-deception to convince herself that her actions won't hurt others and, as a result, allow herself to take the selfish action. Relatedly, Konow (2000) studies behavior in a variant of the dictator game where the size of the pie to be split by the dictator was determined by effort choices made in the past by both the dictator and the recipient. This feature creates the possibility of multiple fairness ideals: e.g., the ideal could be to split the pie in half or, alternatively, to allocate the pie proportional to effort. Konow (2000) argues that a dictator can allow herself to be more selfish by choosing the fairness ideal that gives her the highest payoff. Two pieces of evidence are particularly relevant to our paper. First, Konow (2000) conducted a spectator treatment in which a third party, who was paid a fixed sum, had to split the money between the two subjects. Consistent with a self-serving bias, dictators in the standard game took for themselves more than what spectators gave to similarly situated subjects. Second, when a dictator had to play as a spectator in a subsequent game, she chose for others the same fairness ideal that she initially had found convenient for herself, suggesting that cognitive dissonance may have had a long-lasting effect on the dictator's choice of fairness ideal.

Our paper is also related to a number of studies that suggest people value strategies which allow them to avoid other-regarding behavior. First, Dana, Weber, and Kuang (2007) argues that individuals may actively avoid information to reduce altruism. In their experiment, subjects were given a choice of either clicking a button on the computer to learn about the effect of their actions on others' earnings or to remain uninformed. Despite the minimal cost of this mechanism, one-half of the participants in their experiment chose strategic ignorance and, subsequently, went on to reduce other-regarding behavior. In contrast, when the choice to remain uninformed was not available, behavior was more altruistic.¹⁰ Second, Hamman, Loewenstein, and Weber (2010) argues that individuals may delegate some choices in order to avoid taking direct responsibility for selfish behavior. They conducted experiments in which principals could either decide how much money to share with a recipient or hire agents to make sharing decisions on their behalf. They show that recipients receive significantly less, and in many cases close to nothing, when the allocation decisions can be delegated.

Third, Lazear, Malmendier, and Weber (2012) provides evidence from a laboratory experiment suggesting that some individuals actively try to avoid having the option to be altruistic toward others, even though they choose to be altruistic when that is an option. DellaVigna, List, and Malmendier (2012) provides related evidence by conducting a field experiment that placed flyers on the doorknobs of houses that were going to be visited by a charity fund-raiser. They show that randomly

the false consensus hypothesis, where those who are more selfish also tend to believe others are more selfish. This hypothesis, however, does not predict differences in beliefs across our two treatments.

¹⁰For work where candidates for dictator prefer to opt out for a fixed fee that is lower than the dictator endowment, see Dana, Cain, and Dawes (2006): see also Oberholzer-Gee and Eichenberger (2008). Haisley and Weber (2010) demonstrates how dictators are more likely to choose an unfair option when the recipient's allocation depends on an ambiguous lottery than on a lottery with a known probability.

providing individuals the option to avoid the opportunity to give, by adding a “Do Not Disturb” box in the flyer, reduced giving by 30 percent.

II. Beliefs and Choices in the Basic Corruption Game

A. *Experimental Design*

The experiment was programmed and conducted in z-Tree (Fischbacher 1999). Subjects entered the lab in groups of 16. We preserved anonymity by paying subjects using random identification numbers.¹¹ Before starting, subjects had to read and sign a consent form explaining that the experimenters were not going to deceive them in any way, that they were making anonymous decisions, and that their choices were actually going to affect their payments and those of their circumstantial partners. Subjects were again reminded about these rules at the start of the experiment through a set of on-screen instructions.

Each subject was then asked to complete five tasks consisting of finding a sequence of binary numbers within a longer sequence. Each task took on average one minute to complete. After completing the five tasks, each player was told that she had earned ten tokens. The fact that subjects earned the same number of tokens is not crucial and it is only meant to simplify the experimental setting by providing equal entitlements.

After working for the tokens, subjects went through a set of detailed instructions explaining how the game worked. Each subject was (anonymously) matched with another subject from the lab and their tokens pooled so that each pair of subjects had 20 tokens. One of the two subjects was randomly assigned the role of allocator and the other subject the role of seller. The allocator’s task was to decide how to split the 20 tokens (between herself and the seller that she was matched to), while the seller’s task was to choose the price at which tokens would be sold to the experimenter. If the seller chose “Option A” then the price of each token was 2 Argentine pesos (i.e., *both* the seller and the allocator were paid AR\$2 per token). If the seller chose “Option B” then the price of each token was AR\$1 (i.e., both the seller and the allocator were paid AR\$1 per token), but the seller got an additional payment of AR\$10 only for her. The actions were simultaneous, so the seller did not know how the allocator split the 20 tokens when choosing option A or B, and the allocator did not know what price the seller chose when choosing how to allocate the tokens. Note that a purely selfish allocator should always prefer to take all 10 tokens from the seller’s pile, no matter what the seller chose. Similarly, the seller earns more cash opting for Option B (except in the case of a very generous allocator who allocates for himself less than one-half of the tokens, which did not happen in this experiment).

Before making their decisions on how to allocate and sell, all subjects had to complete a questionnaire about the rules of the game. In order to give them incentives to pay attention to the rules, they were told in advance that we would pay them extra for each correct answer. There were four questions. In the first two, subjects

¹¹ We took the subjects’ name when they entered the lab but then we let them choose where to sit (there were 16 computers available at any one time). In front of each computer there was a number and subjects were told that they would need that number at a later date to recover an envelope with their payment.

were given hypothetical choices for both allocator and seller, and they had to calculate the resulting payments for both players.¹² The answers of the subjects were correct over 70 percent of the time. The last two items of the questionnaire were two statements which the subject had to determine were true or false, with 85 percent (95 percent) of subjects answering the first one (second one) correctly.¹³ In total, 43 percent answered all questions right and 33 percent answered three questions correctly. After answering each question, subjects faced a screen indicating whether they had selected the right answer and a detailed explanation on how to get there (even if the answer they entered was the right one).

Allocators then proceeded to the stage where they had to split the 20 tokens. They faced a screen with two rectangular areas representing the “box” of the allocator and that of the seller. Ten circular tokens were on the allocator’s side and the other ten were on the seller’s side. The allocator can transfer the yellow tokens between the two players with a click-drag-and-drop of the mouse. A subset of tokens, painted in green, is “blocked” (i.e., they cannot be moved by the allocators). For an English translation of all the instructions, including a screenshot of the allocator’s interface for transferring tokens, see the online Appendix. We randomized allocators into two treatments: $Able=8$ (i.e., the subject is able to move up to 8 tokens) and $Able=2$ (i.e., the subject can only move up to 2 tokens).

The key aspect of our design is that the sellers were kept blind about the fact that a certain number of tokens are blocked. The allocators are told that the sellers do not know how the tokens are distributed before making their decision. Moreover, allocators are presented with *all* the instructions given to the sellers, which do not contain any reference to the blocked tokens. As a result, the allocators should expect sellers to believe that $Able=10$. Allocators are not told whether other allocators in the lab may face different values of $Able$.

Given this structure of information, an allocator that only cares about material consumption should form beliefs independently of the value of $Able$. On the other hand, if the allocator cares about her image, she would have found it useful to engage in self-deception (Hypothesis 1), and we should observe that those with a higher $Able$ (i.e., those who could take more tokens) are more likely to convince themselves that sellers chose the unkind action (Option B).

We retrieved two measures of the allocators’ beliefs. First, we asked allocators whether they thought the particular seller to which they were matched chose Option A or Option B. The variable *Is Corrupt* takes the value 0 if the allocator answered “Option A” and 1 if the allocator answered “Option B.” There was no monetary reward for making the correct guess in this version of the experiment, as we wanted to give subjects an opportunity to express their beliefs without a cost. The allocator was also asked to explain (anonymously) her answer on a piece of paper. The goal was to ensure that she had the opportunity to think in more detail about his choice and about the seller. In the following screen, the allocator was given a bonus question: “What percentage of sellers playing today in the lab chose Option B?”

¹²In the first hypothetical situation the allocator keeps 10 tokens and the seller chooses B. In the second, the allocator keeps 19 tokens and the seller chooses A.

¹³The questions were: “The other players or the experimenter will be able to identify your decisions in the game”; and “Even though they do not know your name, the seller knows how you split the tokens at the time of choosing A or B.” The correct answer is *False* in both cases.

There were ten possible answers and we constructed the variable *%-Corrupt* accordingly: if the subject chose the category “0–10%” we assigned it 0, whereas if they chose “10–20%” we assigned it 0.1, and so on. To give incentives for truth-telling we chose a substantial reward: the allocator was told that he would be awarded AR\$20 if her answer was correct, which amounts to over 60 percent of the average final payments received during the experiment.

Given that subjects were randomly assigned to the two treatments, we can directly compare the distribution of beliefs about the seller across the two groups of allocators (those with *Able*=8 and those with *Able*=2). By comparing the effect of the treatment on the allocator’s belief about the seller, we are testing Hypothesis 1.

When asked about the actions of sellers, one concern is that individuals could provide reports with a view to justify their own behavior to the experimenter. Indeed, there is evidence that individuals sacrifice consumption to appear altruistic to others: e.g., giving anonymity to the dictator leads to less altruistic behavior (Hoffman, McCabe, and Smith 1996; Andreoni and Bernheim 2009).¹⁴ We took three steps to address this concern. First, we used a blind design so that a subject (or the experimenter) could not match the choices they observe with the identity of the people in the lab. We emphasized the fact that there was anonymity in the consent form, during the instructions, and in the questionnaire. Second, we provided significant monetary incentives to report truthful beliefs in the form of a large reward for an accurate guess about the proportion of sellers who took the unkind action. Third, in order to be able to compare answers elicited under different incentives for accuracy, we did not provide rewards for accurately guessing the action of the particular seller with whom they were matched.

After finishing the game, all the subjects took an on-screen survey that collected basic information such as gender, age, and socioeconomic status.

The experiment took place at a leading private university in Argentina. Participants were drawn from a database of students who declared to be interested in participating in experiments. Students were not informed of the content, only that it would take place in front of a computer, that participants would be asked to perform simple tasks, that the decisions in the experiment were anonymous, and that they could earn some money in return. Most of the students in this university belong to families in the highest decile of the income distribution of Argentina.

Subjects were permitted to participate only once. There was no show-up fee. The subjects earned on average AR\$38 (just under US\$10 at the time), and the time spent since they entered the lab until they left was around 30 minutes. The stakes were reasonably high: a student working for the university (e.g., in the library) was typically paid up to US\$4 per hour (although opportunities for work were very limited). Consistent with this, all the subjects reported that they would like to be called for a future experiment.

We employed a total of 64 subjects (i.e., 32 allocators and 32 sellers) split in four sessions. The choices made by the sellers are not relevant for the analysis, so in the remainder of the paper all data correspond to the allocators. We note that 75 percent

¹⁴Even in such a case, and given that it does affect altruistic behavior, it would be important to understand to what extent individuals exploit ambiguity to justify their behavior to others. Andreoni and Sanchez (2014) provides evidence showing that individuals report untruthful beliefs to appear fair.

TABLE 1—DATA DEFINITIONS

Name	Definition
<i>Is Corrupt</i>	Dummy variable that takes value 0 if the allocator guessed that her corresponding seller chose Option A, and 1 for Option B.
<i>%-Corrupt</i>	“What percentage of sellers playing today in the lab chose Option B? 0–10 percent (0); 10–20 percent (0.1); ...; 90–100 percent (0.9).”
<i>Able=8</i>	Dummy variable that takes value 1 if the allocator could take up to 8 tokens from the seller’s pile (i.e., faced 2 blocked tokens), and 0 if she could take up to 2 tokens from the seller’s pile (i.e., faced 8 blocked tokens).
<i>Tokens Taken</i>	Number of tokens taken from the seller by the allocator.
<i>Socioeconomic class</i>	“What is the socioeconomic class of your family? Lower class (1); Middle-lower class (2); Middle class (3); Middle-higher class (4); Higher class (5).”

TABLE 2—SUMMARY STATISTICS IN THE FOUR SAMPLES (*For allocators*)

Experiment	First subject pool		Second subject pool	
	Basic game	Modified game	Forced seller	Forced allocator
<i>Is Corrupt</i>	0.67 (0.48)	0.68 (0.47)		0.53 (0.51)
<i>%-Corrupt</i>	0.59 (0.31)	0.56 (0.23)		0.56 (0.24)
<i>Able=8</i>	0.50 (0.51)	0.52 (0.50)	0.53 (0.50)	
<i>Tokens Taken</i>	3.77 (3.05)	4.09 (3.03)	2.66 (3.35)	5.00 (3.05)
<i>Female</i>	0.53 (0.51)	0.44 (0.50)	0.31 (0.46)	0.53 (0.51)
<i>Age</i>	21.07 (2.20)	18.79 (0.92)	20.69 (1.18)	21.33 (3.25)
<i>Socioeconomic class</i>	3.50 (0.63)	3.02 (0.57)	2.92 (0.68)	3.03 (0.56)
Observations	30	65	59	30

Notes: Average characteristics, with standard deviations in parentheses. Observations corresponding to the allocators only.

of sellers choose Option B (low price + side payment) over Option A. We had to discard two allocators, but including them does not alter any of the results below.¹⁵ Variable definitions appear in Table 1, and their corresponding descriptive statistics in Table 2.

¹⁵One of the subjects declared at the end that he had not understood the rules of the game. As a confirmation, we note that he spent over 15 minutes solving the last of the initial 5 tasks (we recorded the time that each subject spent on each screen of the experiment). Therefore, we conclude that he did not understand the rules not because they were difficult to understand, but because he had to rush over the rest of the game to compensate for the time lost. We discarded the second observation because he declared in the questionnaire that he was not a student from the university.

TABLE 3—CHOICES, BELIEFS, AND CHARACTERISTICS IN THE FOUR GAMES, BY TREATMENT GROUP

	First subject pool			Second subject pool					
	<i>Panel A. Basic game</i>			<i>Panel B. Modified game</i>					
	Able=2	Able=8	<i>p</i> -value*	Able=2	Able=8	<i>p</i> -value*			
<i>Posttreatment</i>									
<i>Tokens Taken</i>	1.53 (0.83)	6.00 (2.80)	<0.01	1.35 (0.88)	6.59 (1.91)	<0.01			
<i>Is Corrupt</i>	0.47 (0.52)	0.87 (0.35)	0.02	0.48 (0.51)	0.85 (0.36)	<0.01			
<i>%-Corrupt</i>	0.49 (0.35)	0.69 (0.24)	0.08	0.48 (0.23)	0.64 (0.20)	<0.01			
<i>Pretreatment</i>									
<i>Female</i>	0.47 (0.52)	0.60 (0.51)	0.48	0.37 (0.49)	0.50 (0.51)	0.31			
<i>Age</i>	21.10 (1.28)	21.10 (2.89)	1.00	18.97 (0.98)	18.65 (0.85)	0.21			
<i>Socioeconomic class</i>	3.47 (0.52)	3.53 (0.74)	0.78	3.00 (0.58)	3.03 (0.58)	0.85			
Observations	15	15	30	34	31	65			
	Second subject pool								
	<i>Panel C. Forced seller</i>			<i>Panel D. Forced allocator</i>					
	Able=2	Able=8	<i>p</i> -value*	Sell A	Sell B	<i>p</i> -value*	Take 2	Take 8	<i>p</i> -value*
<i>Posttreatment</i>									
<i>Tokens Taken</i>	1.07 (1.12)	4.10 (4.00)	<0.01	2.44 (2.93)	2.84 (3.70)	0.65	2.00 (0.00)	8.00 (0.00)	0.00
<i>Is Corrupt</i>							0.53 (0.52)	0.53 (0.52)	1.00
<i>%-Corrupt</i>							0.55 (0.24)	0.56 (0.24)	0.94
<i>Pretreatment</i>									
<i>Female</i>	0.25 (0.44)	0.35 (0.49)	0.39	0.37 (0.49)	0.25 (0.43)	0.33	0.47 (0.52)	0.60 (0.51)	0.48
<i>Age</i>	20.54 (0.64)	20.84 (1.51)	0.33	20.59 (1.05)	20.78 (1.29)	0.55	21.20 (3.63)	21.47 (2.95)	0.81
<i>Socioeconomic class</i>	2.82 (0.61)	3.00 (0.73)	0.32	2.96 (0.65)	2.88 (0.71)	0.62	3.20 (0.68)	2.87 (0.35)	0.10
Observations	28	31	59	27	32	59	15	15	30

Notes: Average characteristics, with standard deviations in parentheses. *p*-value of the standard mean difference test whose null hypothesis is that the mean of the distributions under *Able=8* and *Able=2* are equal. Observations corresponding to the allocators only.

B. Results of the Basic Game

In panel A of Table 3 (First Subject Pool, Basic Game), we compare the mean values of key variables across allocators randomly treated with *Able=8* and with *Able=2*. First, note that differences in “pretreatment” characteristics such as gender, age, and socioeconomic class are not statistically significant, suggesting that treatment assignment was effectively random.¹⁶ The data also show that allocators with

¹⁶ As an additional robustness check, we cannot reject equality of the proportion of sellers choosing option A across treatment groups.

$Able=8$ took substantially more tokens from the sellers than allocators with $Able=2$. This confirms that, as expected, the restriction to take up to 2 tokens was binding for a majority of allocators.

We can now test the main hypothesis of the paper. Under the null hypothesis of no self-serving bias, the two beliefs (*Is Corrupt* and *%-Corrupt*) should not be affected by the *Able* treatment. Table 3 shows that allocators with $Able=8$ reported a belief *Is Corrupt* that is 40 percentage points greater, and a belief *%-Corrupt* that is 20 percentage points greater than allocators treated with $Able=2$. These differences are statistically significant, with p -values of 0.019 and 0.078, respectively. In terms of magnitude, *Is Corrupt* increases by 85 percent (from 0.47 to 0.87), while *%-Corrupt* increases by 41 percent (from 0.49 to 0.69). Note that the effects on these two variables should not be compared to each other, because while *%-Corrupt* corresponds to a probability belief, *Is Corrupt* is an indicator variable for the most likely scenario.¹⁷

Figure 1 presents these same results in graphical form. Panel A of Figure 1 shows the distribution of the number of tokens taken, while panel B shows the distributions of *%-Corrupt*, both by treatment group. Panel A shows that $Able=8$ allocators take more tokens than those in the $Able=2$ treatment. In turn, panel B shows that $Able=8$ allocators are more likely to believe that sellers choose the unkind action (low price + side payment).

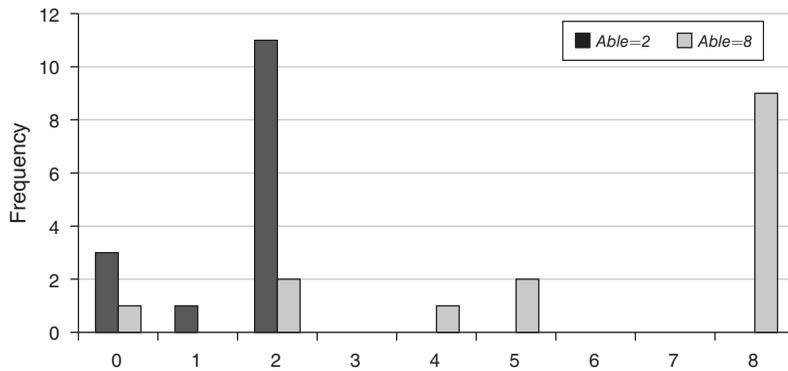
In Table 4 we present the regression-equivalent of the mean difference tests reported in panel A of Table 3. The top half of the table shows the effects of $Able=8$ (relative to $Able=2$) on *Tokens Taken*, *Is Corrupt*, and *%-Corrupt*. For example, the number 4.467 in column 1 is the estimated coefficient in a regression of *Tokens Taken* on a dummy variable that takes the value 1 if $Able=8$ (and zero if $Able=2$). By construction, the results in column 1 are the same than the results shown in Table 3. Column 2 shows that the results are very similar after including demographic variables as additional controls.

We can assess the quantitative importance of the bias identified here. The first approach is to estimate how much of the variation in beliefs about the seller is due to cognitive dissonance. Assuming subjects in the $Able=2$ treatment did not experience cognitive dissonance, then the difference in *%-Corrupt* between $Able=8$ and $Able=2$ subjects would give a measure of the self-serving bias. Thus, 29 percent = $(0.69 - 0.49)/0.69$ of the $Able=8$ seller's beliefs is due to cognitive dissonance. In reality, however, $Able=2$ subjects also likely experienced cognitive dissonance, so this estimate provides a lower bound to the magnitude of bias.

To provide a more precise measurement, the last two rows from Table 4 provide a measure of belief distortion per each token taken. To derive this, we estimated an instrumental-variable regression where the dependent variable is the belief about the seller, the endogenous variable is the number of tokens taken by the allocator, and the instrumental variable is the treatment dummy (i.e., $Able=8$ versus $Able=2$).

¹⁷To illustrate, consider the case where all subjects with $Able=2$ have the same value for *%-Corrupt*, equal to 0.5. As a result, one-half of them choose *Is Corrupt* equal to 1 and one-half choose 0. If $Able=8$ increases *%-Corrupt* by 0.1, the average *Is Corrupt* will increase to 1, because all subjects will now think that the seller being corrupt is the most likely scenario. The finding that the effect on *Is Corrupt* is approximately twice as large as the effect on *%-Corrupt* turns out to hold also in the Modified Game (where subjects are rewarded for the accuracy in both beliefs).

Panel A. Tokens taken by the allocator



Panel B. Allocator's belief about the seller (%-Corrupt)

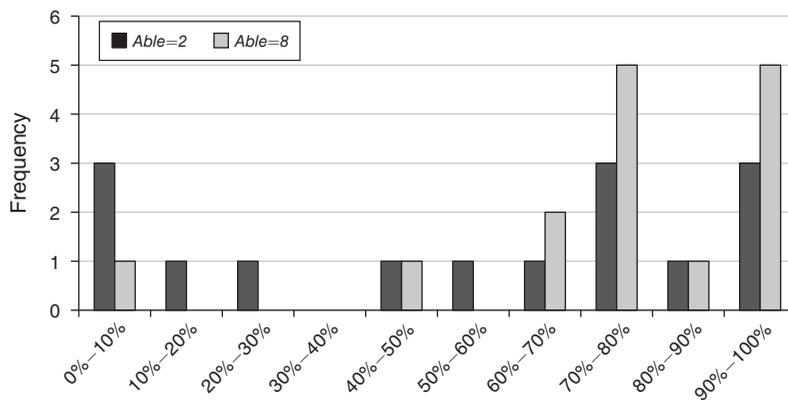


FIGURE 1. DISTRIBUTION OF ACTIONS AND BELIEFS BY TREATMENT GROUP, BASIC GAME

According to these estimates, for each extra token taken by the allocator, cognitive dissonance increases %-Corrupt by 0.045 (last row from column 1 of Table 4). Since individuals in the Able=8 group take on average 6 tokens, this suggests that cognitive dissonance accounts for $0.27 = 0.045 \times 6$ of the belief %-Corrupt in the Able=8 group, or 39 percent = $0.27/0.69$ of the average belief in that group. Individuals in the Able=2 group do not have as much cognitive dissonance because they can take fewer tokens. Since they take on average 1.53 tokens, that implies that cognitive dissonance accounts for 14 percent = $(0.045 \times 1.53)/0.49$ of the average belief about the seller in the group Able=2.

Below we measure whether, as predicted by Hypothesis 2, the ability to engage in self-serving beliefs affected the allocator's decision to take tokens from the seller. This will provide an alternative way of quantifying the implications of cognitive dissonance.

TABLE 4—SUMMARY OF TREATMENT EFFECTS IN THE BASIC AND MODIFIED GAMES

	Basic game		Modified game	
	(1)	(2)	(3)	(4)
<i>Effect of Able=8 (versus Able=2) on Tokens Taken</i>	4.467*** (0.755)	4.293*** (0.603)	5.233*** (0.363)	5.150*** (0.402)
<i>Is Corrupt</i>	0.400** (0.161)	0.344*** (0.160)	0.369*** (0.110)	0.325*** (0.120)
<i>%-Corrupt</i>	0.200* (0.110)	0.182 (0.111)	0.158*** (0.053)	0.151*** (0.055)
<i>Implied effect of Tokens Taken on Is Corrupt</i>	0.090** (0.032)	0.080** (0.033)	0.071*** (0.021)	0.063** (0.024)
<i>%-Corrupt</i>	0.045* (0.024)	0.042 (0.026)	0.030*** (0.010)	0.029*** (0.010)
Controls	No	Yes	No	Yes
Observations	30	30	65	65

Notes: Heteroskedastic-robust standard errors in parentheses. All regressions are ordinary least squares. The control variables are: gender, age, and socioeconomic class. The Implied Effect of Tokens Taken is the coefficient from an instrumental variable regression where the dependent variable is the belief about the seller, the endogenous variable is the number of tokens taken, and the instrumental variable is the dummy for *Able=8*. Observations corresponding to the allocators only.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

III. Further Evidence from Three Additional Games

In this section we present three additional experiments designed to study the boundaries of the main effect. In Section IIIA we present the “Modified Game,” intended to study the robustness of the results in the Basic Game to several modifications in the design of the experiment. In Section IIIB we present the Forced Seller Game, designed to test whether the ability to engage in self-deception affects the allocator’s choice to be selfish (i.e., Hypothesis 2). Finally, in Section IIIC we present the Forced Allocator Game, designed as a falsification test for the main results.

We conducted these experiments with a different subject pool, recruited from the undergraduate population at a (different) private university. Subjects were enrolled in large, required courses in economics and business administration, and were recruited through verbal announcements after classes. There were seven sessions, and subjects were permitted to participate only once. As Table 2 suggests, the characteristics of participants from this new subject pool were similar to those of subjects participating in the Basic Game in terms of age and gender, but from a somewhat less privileged economic background.

As in the Basic Game, subjects had to read and sign the consent form. There was no show-up fee. Subjects earned on average AR\$24.70 (approximately US\$3.20 at the time), and the amount of time spent on the experiment was around 30 minutes. The stakes were reasonable: at that university, students who had access to work

(e.g., in the library) could earn approximately US\$1 per hour (although work opportunities were very limited). Three-quarters of subjects reported that they would like to participate in a future experiment, which is somewhat lower than in the Basic Game, but that is not surprising given the differences in the recruitment process.

A. *The Modified Corruption Game*

Design of the Modified Corruption Game.—The “Modified Corruption Game” was designed to study the robustness of the basic results to several changes in the design of the experiment and, additionally, to changing the pool of subjects. The design is identical to the Basic Game, with the following exceptions:

Paper Format: Subjects were given a paper version of the instructions (in contrast to the Basic Game, where they conducted the experiment in computers). This change was made primarily to facilitate the implementation of the experiment with a larger number of subjects.

Questionnaire about the Rules of the Game: In the Basic Game, when subjects answered a question about the design of the game, we had the option of showing them the correct answer on the screen, together with a detailed explanation. This was not easy to implement in a paper-and-pencil context. To provide incentives to pay attention, we told subjects we would not pay them anything unless all questions about the rules of the game were answered correctly. Over 80 percent of subjects answered correctly all questions. We do not include data for subjects who answered one or more questions incorrectly (but we note that the results are robust to the inclusion of these observations).

Explicit Randomization: Subjects were explicitly told that there were two values of *Able*, namely 2 and 8, and that subjects were randomly assigned to one of the two groups with equal probability. They were then told to which of the two groups they had been assigned. This variation in the design is helpful in the event subjects in the Basic Game did not fully understand the fact that sellers were unaware of the existence of the treatment. This manipulation also serves to refute a possible confound: it is possible that allocators in the Basic Game “learned” something from the particular value of *Able* they faced. For example, a low value of *Able* could have been interpreted by the allocator as a signal that sellers behaved generously in the past and, thus, a signal of the behavior that was expected of them (e.g., that allocators should not take many tokens). The key aspect of our design was maintained: allocators are shown all the instructions given to the sellers, which explicitly note that sellers do not know if the particular allocator to whom they were matched was assigned to $Able=8$ or $Able=2$.

Seller’s Options: If the seller chose “Option A” then the price was AR\$1.50 per token (i.e., both the seller and the allocator got AR\$1.50 per token) while if he chose “Option B” then the price per token was AR\$0.50 (i.e., both the seller and allocator got AR\$0.50 per token), but the seller got an additional payment of AR\$5 only for herself. Selling at the low price + side payment was no longer a preferred strategy

(for allocators expected to take positive amounts of tokens). Instead, sellers that expect the allocator to take four or fewer tokens were better off selling at the high price. This was done to make the environment more conducive to trusting sellers.

Belief Elicitation: The order of the belief questions was changed: allocators were first asked whether the seller they were matched with took Option B, then took their action, and finally responded the proportion of sellers that took the unkind action. Also, the two beliefs were elicited on the same page where the decision to allocate tokens was made. This was done to ensure that subjects knew there was a cost to holding inaccurate beliefs *before* they made their choice. Additionally, we offered AR\$5 incentives for eliciting each of the two beliefs. Thus, while only %-Corrupt was rewarded in the Basic Game, we now rewarded both the *Is Corrupt* and %-Corrupt beliefs. If beliefs do not respond to the size of incentives, it suggests that beliefs are really formed by actions (and may influence behavior more broadly).

Anonymity: We now recruited students immediately after class so they never had to provide their names to the experimenters. They just decided to take part in an experiment, were given a piece of paper with a number, and were told they would have to come back at a later date to pick up their payment in an envelope with that number. To ensure understanding, subjects are asked, “Can the organizers of the experiment know your identity?” Only subjects who answered correctly (“no”) are used (only 4 out of the original 81 allocators got this question wrong).

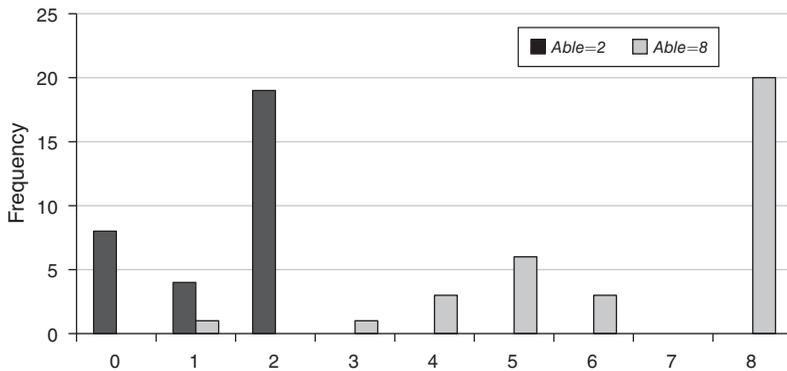
Results of the Modified Corruption Game.— We employed a total of 130 subjects (i.e., 65 allocators and 65 sellers). As before, we show data for the allocators only, noting simply that 66 percent of sellers choose Option B (low price + side payment) over Option A.

In panel B of Table 3 of the Second Subject Pool, we compare choices, beliefs, and other variables across the two treatment groups for the Modified Game. There are no significant differences in pretreatment characteristics, which indicates that treatment assignment was effectively random.¹⁸ Allocators with *Able*=2 took on average 1.35 tokens from the sellers, while those with *Able*=8 took on average 6.6 tokens, with the difference being statistically significant. Both measures of beliefs are significantly different across treatments, and the magnitude of these differences appears similar to that observed in the Basic Game. Panels A and B of Figure 2 show the distribution of *Tokens Taken* and %-Corrupt, both by treatment group. As in the Basic Game, the raw data suggest that allocators in the *Able*=8 treatment take more tokens and are more likely to believe sellers chose Option B.

In Table 4, columns 3 and 4, we present the regression results for the Modified Game. The results are remarkably similar to those obtained in the Basic Game. In the top half we present regression estimates of the mean difference tests in Table 2, again confirming that the treatment was effective in causing allocators to take more tokens. The effect of *Able*=8 on both measures of beliefs is statistically significant at the 1 percent level, both with and without control variables. The estimated effects

¹⁸ Similarly, we cannot reject equality of the proportion of sellers choosing option A across treatment groups.

Panel A. Tokens taken by the allocator



Panel B. Allocator's belief about the seller (%-Corrupt)

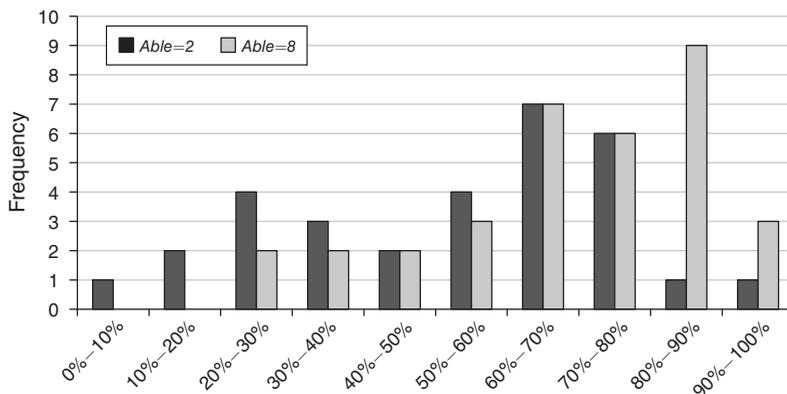


FIGURE 2. DISTRIBUTION OF ACTIONS AND BELIEFS BY TREATMENT GROUP, MODIFIED GAME

on beliefs appear to be somewhat smaller in the Modified Game than in the Basic Game, but these differences are not statistically significant.

Finally, the bottom half of Table 4 presents the estimates from the instrumental variables regression where the dependent variable is the belief about the seller, the endogenous variable is the number of tokens taken, and the instrumental variable is the dummy for the treatment. As in the Basic Game, we can use these estimates to measure how much of the variation in beliefs is due to cognitive dissonance. According to these estimates, for each extra token taken by the allocator, cognitive dissonance increases %-Corrupt by 0.03. Since individuals in the Able=8 group take on average 6.6 tokens, the self-serving bias increases the belief about the proportion of sellers taking the unkind action by 20 percentage points (i.e., 0.03×6.6). This is our “preferred” specification, which we use as one measure of the magnitude of the self-serving bias. This 20-percentage-points effect accounts for 31 percent = $(0.03 \times 6.6)/0.64$ of the average belief about the sellers in the Able=8

condition. This is close to the corresponding estimate from the Basic Game (which is 39 percent; the difference is not statistically significant).

The similarity in results across the Basic Game and the Modified Game suggests that letting subjects know that accuracy in beliefs is profitable before they have to make their decisions in the game does not reduce significantly the extent of the bias. One possible explanation is that the allocator starts forming beliefs about the sellers immediately after reading the instructions, so that by the time when she finds out about the existence of a reward for accurate beliefs she already incurred in self-deception. An alternative explanation is that cognitive dissonance does not internalize all the costs and benefits from holding beliefs, such as rewards from accurate beliefs. See the model in the Appendix for a discussion.

B. *The Forced Seller Game*

Design of the Forced Seller Game.—In order to test Hypothesis 2, namely that the ability to engage in self-deception affects the allocator's choices, we conducted a new experiment which we call the Forced Seller Game. Variables are identical to the Modified Game, with one critical difference: a computer makes the choice on behalf of the seller. Allocators were randomized into two groups: one that would face a seller who was "forced" to sell at a high price (Option A), and another group who would face a seller who was "forced" to sell at the low price plus the side payment (Option B). As before, the allocator could see the instructions given to the particular seller to whom she was matched and thus knew the seller was forced to sell at a predetermined price. As in the Modified Game, allocators knew that they had been randomly assigned to one of the two values of *Able*. But allocators did not know that there was another group of allocators facing a different type of sellers.

This game is almost identical to the standard dictator game as it introduces a passive recipient, fully eliminating the ambiguity regarding the seller's actions and her altruism. As a result, allocators can no longer bias their beliefs about the possible action taken by the sellers (and hence bias their inference regarding the seller's altruism). Even though in this setting we cannot study beliefs about the sellers' actions (since these are known with certainty), we can still study the allocator's actions (i.e., *Tokens Taken*). This allows us to test Hypothesis 2: the ability to manipulate beliefs about others induces people to behave more selfishly. Under Hypothesis 2, eliminating the ability to engage in belief manipulation should reduce the number of tokens taken by allocators (relative to the Modified Game).

Results of the Forced Seller Game.—The average number of tokens taken in the Forced Seller game, reported in Table 2, is 2.66. This is lower than the average number of tokens taken in the Modified Game, which amounts to 4.09 (the difference is statistically significant at the 1 percent level). In other words, eliminating ambiguity regarding the seller's actions significantly reduced the number of tokens taken. Or, put differently, the evidence suggests that ambiguity about others' social preferences increases the likelihood that a person will be selfish.

One way to establish the importance of a self-serving bias in beliefs is to quantify its effect on the allocators' choice (i.e., the number of tokens taken). We approach this through two different strategies. The first, more direct, estimate compares the

outcomes in the Modified Game and the Forced Seller game.¹⁹ Table 3 presents the average number of tokens taken by treatment ($Able=8$ and $Able=2$). In the Modified Game, allocators in the $Able=8$ treatment took on average 6.59 tokens. In the Forced Seller, allocators in the $Able=8$ treatment took on average 4.10 tokens. That is, eliminating the ambiguity about the sellers' intention (which eliminates the scope for self-deception) is estimated to reduce the average number of tokens taken by the allocator in the $Able=8$ condition by 2.49 (standard error 0.48, p -value < 0.01). This would amount to 25 percent of the total endowment of the seller (10 tokens), and 38 percent of the average number of tokens taken in the $Able=8$ treatment (6.6 tokens).

This estimate, along with the corresponding effect on beliefs reported for the Modified Game, provides our preferred measure of the importance of biased beliefs: i.e., cognitive dissonance increases the allocator's belief regarding the proportion of sellers that took the "unkind" action by 20 percentage points, and makes the allocator take an additional 2.5 (out of 10) tokens from the seller.

As a robustness check, we provide a second, indirect, measure of the effect of cognitive dissonance on the allocator's actions, which does not use data from the Forced Seller Game. As stated above, our preferred estimates suggest that cognitive dissonance increases $\%Corrupt$ by 0.20. We need to estimate how many extra tokens the allocator wants to take because of this change in beliefs. As a first approximation, we estimated an interval regression of the number of tokens taken on $\%Corrupt$ using the data from the Modified Game.²⁰ The regression coefficient indicates that a 0.1 increase in $\%Corrupt$ increases the tokens taken by the allocator by 0.96 (standard error 0.43, p -value = 0.025). This correlation suggests that the 0.20 increase in $\%Corrupt$ due to cognitive dissonance would allow an allocator to take up to $1.92 = 0.2 \times 9.6$ additional tokens (standard error 0.63, p -value < 0.01).²¹ It is interesting that the difference in the two estimates (1.92 and 2.49) is statistically insignificant.

One potential concern with this last calculation is that the coefficient from the regression of the number of tokens taken on $\%Corrupt$ may be subject to omitted-variable biases. For example, the tendency to distrust the seller may be positively correlated to the (unobserved) selfishness of the allocator. As an additional robustness check, we later conducted another modified version of the Modified Game (i.e., a fourth additional game conducted one year later with a sample of first-year students), which we denominate the Conditional Game. The most important change in

¹⁹ As suggested earlier, subjects for the Modified and Forced Seller games are drawn from the same population. Given our interest in preserving anonymity vis-à-vis the experimenter, we did not collect the subjects' basic information to conduct random assignment into these two games. As a result, there are small but statistically significant differences in age (Table 2). In unreported results we find that controlling for the age difference in a number of ways has a negligible effect on the way subjects play these games and the results we report.

²⁰ Ideally we would want to estimate the effect of $\%Corrupt$ on the desired number of tokens to take from the seller's pile. We need to use an interval regression model to account for that fact that allocators cannot always take the desired number of tokens. In other words, we must take into account that, when an allocator in $Able = 2$ takes exactly 2 tokens, we only know that her desired number of tokens is equal or greater than 2.

²¹ This simple estimate of the standard error takes 9.6 as fixed, thereby providing a lower bound. We can reproduce this analysis for the Basic Game. The estimates from the interval regression suggest that a 0.1 increase in $\%Corrupt$ is associated with an additional 0.87 tokens taken by the allocator. In turn, this implies that the 0.27 increase in $\%Corrupt$ due to cognitive dissonance could translate in $2.35 = 0.27 \times 8.7$ additional tokens taken from the seller.

this new game is that allocators can make their decision conditional on *%-Corrupt*.²² In other words, allocators must report one allocation decision for each of the ten possible scenarios (i.e., *%-Corrupt* = 0–10 percent, *%-Corrupt* = 10–20 percent, etc.), and we promised them that we would execute the one choice that corresponds to the actual value of *%-Corrupt* in the experiment. With these data, we can measure the causal effect of *%-Corrupt* on the number of tokens taken from the seller. We obtained a sample of 50 subjects (25 allocators) and find that a 0.1 increase in *%-Corrupt* increases the desired number of tokens taken from the seller by 0.62 (standard error 0.198, p -value < 0.01).²³ Even though this 0.62 estimate is smaller than the corresponding estimate of 0.96 reported above, their difference is not statistically significant. This finding supports the validity of the calculations presented above.

Finally, the results from the Forced Seller Game also provide suggestive evidence regarding the reasons why allocators care about the seller's decision. Table 3 presents the average number of tokens taken by the type of seller to which they were assigned. There is a small difference in the mean number of tokens taken across the *Sell=A* and *Sell=B* treatments, but the difference is statistically insignificant.²⁴ That is, even though allocators seem to be sensitive to the price chosen when the seller is making the choice, it seems that allocators do not care about this price when it is chosen by a computer. This finding is consistent with our interpretation that allocators care about the sellers due to feelings of reciprocity. A more provocative way of illustrating the importance of self-serving beliefs is to note that allocators facing the certainty of being "cheated" by a computer are more generous (take only 2.84 tokens: see Table 3) than when they face the possibility of being "cheated" by the seller (they take 4.09 tokens: see Table 2), with the difference being statistically significant.

C. Forced Allocator Game

Design of the Forced Allocator Game.—The Forced Allocator Game is similar to the Modified Game but now the computer makes the choice of how many tokens to take on the allocator's behalf. The instructions explained that allocators would be split into two groups with equal probability.²⁵ One group of allocators would

²²In keeping the structure of the game as close as possible to the Modified Game, the instructions of the seller did not include a reference to the ability of the allocators to make a conditional decision. Additionally, to avoid restricting the variation in the number of tokens taken, we fixed $Able=10$.

²³This estimate corresponds to a Tobit regression of the tokens to be taken by each pair allocator/scenario on the *%-Corrupt* corresponding to the scenario, with censoring points in zero and ten tokens (corresponding to the two corner solutions observed in the data).

²⁴Similarly, previous work has found that subjects in the ultimatum game are significantly less likely to reject a low offer when it was made by a computer as opposed to a real subject (e.g., Blount 1995; Falk, Fehr, and Fischbacher 2008; and Bellemare, Kroger, and van Soest 2011). Note that within each of the two conditions (Seller A, Seller B), there are two treatments ($Able = 8$ and $Able = 2$). The results across each of the four treatments are as follows (average Tokens Taken, standard deviation, number of observations): for allocators in the $Able=8$ /high price (3.5, 3.4, 16), $Able=8$ /low price + side payment (4.7, 4.6, 15), $Able=2$ /high price (0.9, 0.9, 11), and $Able=2$ /low price + side payment (1.2, 1.2, 17).

²⁵Unfortunately, a mistake on our part in one of the sessions of the Forced Allocator Game resulted in fewer usable subjects (15 allocators). The results we report in Table 3 also include 6 allocators who answered some questions wrong, and 9 from another subject pool (from a nearby university). None of the results change if some—or all—of these extra subjects are excluded.

be “forced” to take two tokens from the seller’s pile, while the other group would be “forced” to take eight tokens from the seller’s pile. Each allocator was then told to which of the two groups of “forced allocators” she was randomly assigned. As before, allocators could see the instructions of the sellers, which revealed that the sellers didn’t know the number of tokens that their corresponding allocators were forced to take.

The Forced Allocator game provides a falsification test for the Modified Game: by removing the allocator’s responsibility, we should expect to eliminate the allocator’s need to justify her actions, and thus the allocator should not have incentives to bias her beliefs about the seller. In other words, we expect forced allocators in the two conditions (i.e., forced to take two and eight tokens) to have similar beliefs about the sellers.

Results of the Forced Allocator Game.—Panel C of Table 3 compares beliefs and other variables across treatment groups. There are no significant differences in pretreatment characteristics, which suggests that the treatment assignment was effectively random. More important, the difference in mean beliefs about the seller across *Take=2* and *Take=8* treatments is close to zero and statistically insignificant. In sum, the evidence from the Forced Allocator experiment is consistent with our interpretation that cognitive dissonance drives the bias in beliefs observed in the Basic Game and in the Modified Game.

As discussed above, the beliefs about the seller in the Forced Allocator game should not be affected by cognitive dissonance. As a result, the average *%-Corrupt* in the Forced Allocator game provides a rough estimate of what the average *%-Corrupt* would have been in the Modified Game if allocators had not incurred in self-deception. As shown in Table 2, the average of *%-Corrupt* in the Forced Allocator game is 0.56. In comparison, the results we presented for the Modified Game imply that allocators who do not engage in self-deception would have had an average of *%-Corrupt* of 0.44.²⁶ It is reassuring that the difference between these two estimates is statistically insignificant.

IV. Discussion

In this paper we presented the results from a series of “corruption” games, namely a dictator game where the recipient has the option of reducing the size of the pie in exchange for a side payment. We found evidence consistent with the hypotheses that people’s beliefs about others are affected by their own desire to be selfish and that ambiguity about other people’s actions plays an important role in allowing selfish behavior.

Our results are potentially relevant for work in other fields, including political economy. In a seminal paper, Piketty (1995) studied how economic systems depend on the beliefs that people hold concerning labor market parameters (such as whether effort pays or not). Later work demonstrated that individual beliefs are indeed a

²⁶In the Modified Game, when *Able=8*, *%-Corrupt* was equal to 0.64, of which (see the calculation above) $0.2 = 0.03 \times 6.6$ was due to self-deception. That implies that, without self-deception, *%-Corrupt* would have been $0.44 = 0.64 - 0.2$. By construction, the results are the same for *Able=2*.

strong predictor of a person's demand for redistribution, while a number of papers studied how these beliefs are formed, including the roles of culture (as in Bisin and Verdier 2000 and Giuliano 2007) and the economic environment (as in Di Tella, Galiani, and Scharfrodsky 2007 and Giuliano and Spilimbergo 2014).²⁷ Our evidence emphasizes a new channel through which economic incentives shape beliefs and actions.

Our paper suggests several avenues for future research. First, we provide one setting where self-serving beliefs arise, but it is an open question how long-lasting these effects are, whether they may be equally malleable for other subject pools (e.g., non-students, individuals living in other countries), or whether these effects "spill" over to other settings. For example, being randomly assigned to have the opportunity to be more selfish in the "corruption game" may spill over to behavior in a subsequent trust game, which would suggest that self-serving beliefs could shape economic behavior more broadly. Second, it would be useful to identify other conditions, besides ambiguity, that facilitate distorting beliefs about other people's preferences. For example, our research design could be extended to measure precisely how "convenient" gathering and recollection of data is used as a means for self-deception (see, for example, the over-weighting of convenient signals in Mobius et al. 2014).²⁸ Last, it would be useful to understand the extent to which individuals are aware of the existence of these biases and take them into account, either consciously or unconsciously. For example, it would be informative to measure if sellers are willing to pay to limit the allocator's ability to engage in cognitive dissonance, perhaps by exposing her to objective information.

APPENDIX

A. A Stylized Model²⁹

We begin the model without motivated beliefs. There are two players, denoted by subscripts 1 and 2. Their endowments are denoted x_i , and their consumptions c_i . The first player can transfer an amount t from the second player, although $(1 - \rho)t$ gets lost in the transference, where $\rho \in [0, 1]$ represents the efficiency of the redistribution. "Intrinsic" utility from consumption is given by the function $U(c_1)$ with the standard properties $U'(\cdot) > 0$ and $U''(\cdot) < 0$.

We model fairness concerns in line with the theories on reciprocal altruism (e.g., Levine 1998; Rotemberg 2005). The utility of the first player is a weighted average of

²⁷See also Denzau and North (1994); Bowles (1998); Bénabou and Tirole (2006); Bénabou (2008); and Rotemberg (2014). For evidence, see Fong (2001); Alesina and La Ferrara (2005); Corneo and Gruner (2002); Alesina and Glaeser (2004); and Kuhn (2011). Part of this literature demonstrates that people living in less developed countries also believe the rich are corrupt and that others cannot be trusted (see Di Tella and MacCulloch 2009 and Aghion et al. 2010). There is also work on how the rich tend to underestimate the extent of income inequality (Norton and Ariely 2011) and their own relative income (Cruces, Perez-Truglia, and Tetaz 2013), possibly to justify low redistribution.

²⁸A striking asymmetry is demonstrated in Sharot, Korn, and Dolan (2011), whereby subjects updated their beliefs more in response to information that was better than expected compared to information that was worse. Indeed, the psychology literature on self-deception emphasizes the role of selective collection and recollection of information (e.g., Dunning, Meyerowitz, and Holzberg 1989), and this has been formally incorporated in some economic models (e.g., Rabin 1995; Rabin and Schrag 1999; Compte and Postlewaite 2004; Mobius et al. 2014).

²⁹For more details, as well as an alternative approach, see Di Tella and Perez-Truglia (2010).

her own intrinsic utility and the intrinsic utility of the second player: $U_1 + \lambda(\lambda_2) \cdot U_2$. The weight $\lambda(\cdot)$ is a function of the first player's belief about the second player's unconditional altruism, denoted λ_2 .³⁰ We assume $\lambda'(\cdot) > 0$, so that agents respond like with like. Also, assume that $\lambda(\cdot)$ is positive and concave. The problem of the first player is then

$$\max_{t \in [0, x_2]} U(x_1 + \rho t) + \lambda(\lambda_2) U(x_2 - t).$$

The comparative statics are straightforward. If the first player believes that the second player is more altruistic (i.e., greater λ_2) then she will keep less of the total resources. A marginal increase in efficiency, ρ , will increase the net amount taken from the second player, ρt , but it may lead to an increase or a decrease in t .

We can now introduce motivated beliefs. The second player was randomly chosen from a population of two types: $\lambda_2 \in \{\lambda_2^L, \lambda_2^H\}$. The actual proportion of low type (i.e., the least altruistic) is given by $p_0 \in [0, 1]$. The first player's belief about that proportion is denoted by $p \in [0, 1]$.

For the sake of brevity, we won't model the process of information acquisition and belief formation explicitly. We assume that, if the first player processes the information rationally, her posterior belief will be that $p = p_0$. Alternatively, the first player may form a biased belief by systematically recalling or ignoring the information available to her: for example, instead of correctly remembering a proportion p_0 of low-type, the individual may try to convince herself that the proportion was actually $p > p_0$. We assume that there is a psychological cost for self-serving biases, which is increasing in the size of the bias: $\mu \cdot C(p - p_0)$, with $\mu > 0$, and $C(\cdot)$ strictly convex and strictly increasing in $|p - p_0|$. For example, the selective recalling of information may take extra time and effort, and it may even make the individual feel guilty. If we let the parameter $\mu \rightarrow \infty$, that limiting case would represent the absence of motivated beliefs, because it would never be optimal to incur in self-serving biases.

The problem of the first player becomes

$$\max_{t \in [0, x_2], p \in [0, 1]} U(x_1 + \rho t) + \lambda(\lambda_2^H + p(\lambda_2^L - \lambda_2^H)) U(x_2 - t) - \mu \cdot C(p - p_0).$$

PROPOSITION 1: *If the solution is interior and $\mu < \infty$, then an increase in ρ changes p^* and t^* in the same direction.*

PROOF:

Take the first order condition with regard to p for the interior solution:

$(\lambda_2^L - \lambda_2^H)\lambda'(\lambda_2^H + p^*(\lambda_2^L - \lambda_2^H)) U(x_2 - t^*) - \mu \cdot C'(p^* - p_0) = 0$. And then differentiate with regard to ρ :

$$\frac{dp^*}{d\rho} = \frac{(\lambda_2^L - \lambda_2^H)\lambda'(\lambda_2^H + p^*(\lambda_2^L - \lambda_2^H)) U'(x_2 - t^*)}{(\lambda_2^L - \lambda_2^H)^2 \lambda''(\lambda_2^H + p^*(\lambda_2^L - \lambda_2^H)) U(x_2 - t^*) - \mu \cdot C''(p^* - p_0)} \frac{dt^*}{d\rho},$$

³⁰Given the dictator setting, an asymmetric formulation is sufficient to capture the relevant effects. In more general settings, reciprocal altruism involves $\lambda_i(\lambda_j)$ for $i = 1, 2$ (see Levine 1998 and Rotemberg 2005).

where the denominator in the RHS is negative if the solution is interior (i.e., it is the second derivative of the objective function with respect to p , which has to be negative for the interior solution to be a maximum), and the numerator is negative given the assumptions. As a result, $\frac{dp^*}{d\rho}$ and $\frac{dt^*}{d\rho}$ must have the same sign. ■

If the first player decided to increase t^* after a reduction in ρ , then he must have also increased p^* . Intuitively, given that the first player is taking more resources, she is willing to pay a psychological cost to have more negative beliefs about the second player and therefore alleviate the “psychological cost” of taking extra resources.

To see how we can provide an alternative, more direct, test of the proposition, rearrange the last equation:

$$\frac{\frac{dp^*}{d\rho}}{\frac{dt^*}{d\rho}} = \frac{(\lambda_2^L - \lambda_2^H)\lambda'(\lambda_2^H + p^*(\lambda_2^L - \lambda_2^H))U'(x_2 - t^*)}{(\lambda_2^L - \lambda_2^H)^2\lambda''(\lambda_2^H + p^*(\lambda_2^L - \lambda_2^H))U(x_2 - t^*) - \mu \cdot C''(p^* - p_0)} = \delta,$$

where δ is zero if $\mu \rightarrow \infty$ (i.e., absence of self-serving biases) and positive otherwise. The null hypothesis $\delta = 0$ gives a more direct test of the proposition. It is straightforward to implement this test, exploiting the exogenous variation in ρ that we generated in the experiment. In terms of the corruption game studied in this paper, we can recover δ by simply running a regression of the allocator’s beliefs about the sellers on the number of tokens taken from the seller, using the dummy $Able=8$ as instrumental variable.

B. Discussion

Assume that the first player’s utility depends directly on the accuracy of belief about the other person’s altruism. We model this by an additional term in the utility function: $-\eta \cdot R(p - p_0)$, with $\eta > 0$ and $R(\cdot)$ strictly convex and strictly increasing in $|p - p_0|$. For example, in the corruption game studied in our paper, the allocator could get a reward by correctly guessing the proportion of sellers who took the unkind action.

One of two possibilities may arise. The first is that, at the time of deciding to incur in cognitive dissonance, the individual fully incorporates the potential costs from holding biased beliefs through $-\eta \cdot R(p - p_0)$, in which case she maximizes

$$\max_{t \in [0, x_2], p \in [0, 1]} U(x_1 + \rho t) + \lambda(\lambda_2^H + p(\lambda_2^L - \lambda_2^H))U(x_2 - t) - \mu \cdot C(p - p_0) - \eta \cdot R(p - p_0).$$

The second possibility is that the individual, either consciously or unconsciously, ignores the additional costs from biased beliefs, in which case she maximizes

$$\max_{t \in [0, x_2], p \in [0, 1]} U(x_1 + \rho t) + \lambda(\lambda_2^H + p(\lambda_2^L - \lambda_2^H))U(x_2 - t) - \mu \cdot C(p - p_0).$$

In the evidence from the Modified Game, it seems that subjects were not reacting to the higher incentives for accurate beliefs, $-\eta \cdot R(p - p_0)$. In terms of the model, the subjects are maximizing the last objective function. There are two potential explanations for this phenomenon:

Timing of Belief Formation.—The self-serving bias depends on the order in which the first player learns about the payoffs in the game. In our corruption game, the allocator first learned about the instructions of the game, and it is possible that she formed her beliefs about the seller at that point. When she later finds out about the existence of a reward for accurate beliefs, she is unaware about her self-serving biases and thus has no incentives to revise her beliefs.

A more extreme modification of the experiment could be used to test this extra hypothesis, by reversing the order in which the first player learns about her payoffs in the game and the reward for accuracy. Assume that the first player was shown the instructions of the game purely as a spectator, not to make a decision but for the sole purpose of making a prediction for a reward. In that case, the player would have no incentives to bias the recalling of information, leading to a posterior belief that the proportion of low-type is p_0 . After providing the guess, this individual would be notified that she is actually playing as the first player and has to make a decision. If the timing of the belief formation is important, as conjectured above, then it should be too late for this individual to incur self-serving biases.

Mental Compartments.—The belief manipulation to reduce cognitive dissonance is not the product of rationally ignoring information, but is a mental reflex that appears when the individual is actively considering behavior that affects others. Since the reward for accurate beliefs does not belong to the category of the other-regarding choice, the decision to bias beliefs does not internalize the direct costs from biased beliefs, $-\eta \cdot R(p - p_0)$.

REFERENCES

- Abbink, Klaus, Bernd Irlenbusch, and Elke Renner. 2002. "An Experimental Bribery Game." *Journal of Law, Economics, and Organization* 18 (2): 428–54.
- Aghion, Philippe, Yann Algan, Pierre Cahuc, and Andrei Shleifer. 2010. "Regulation and Distrust." *Quarterly Journal of Economics* 125 (3): 1015–49.
- Akerlof, George A., and William T. Dickens. 1982. "The Economic Consequences of Cognitive Dissonance." *American Economic Review* 72 (3): 307–19.
- Alesina, Alberto, and Eliana La Ferrara. 2005. "Preferences for Redistribution in the Land of Opportunities." *Journal of Public Economics* 89 (5–6): 897–931.
- Alesina, Alberto, and Edward L. Glaeser. 2004. *Fighting Poverty in the US and Europe: A World of Difference*. Oxford: Oxford University Press.
- Andreoni, James, and B. Douglas Bernheim. 2009. "Social Image and the Fifty-Fifty Norm: A Theoretical and Experimental Analysis of Audience Effects." *Econometrica* 77 (5): 1607–36.
- Andreoni, James, and Alison Sanchez. 2014. "Do Beliefs Justify Actions or Do Actions Justify Beliefs? An Experiment on Stated Beliefs, Revealed Beliefs, and Social-Image Manipulation." National Bureau of Economic Research Working Paper 20649.
- Babcock, Linda, and George Loewenstein. 1997. "Explaining Bargaining Impasse: The Role of Self-Serving Biases." *Journal of Economic Perspectives* 11 (1): 109–26.
- Babcock, Linda, Xianghong Wang, and George Loewenstein. 1996. "Choosing the Wrong Pond: Social Comparisons in Negotiations that Reflect a Self-Serving Bias." *Quarterly Journal of Economics* 111 (1): 1–19.

- Bellemare, Charles, Sabine Kroger, and Arthur van Soest.** 2011. "Preferences, Intentions, and Expectation Violations: A Large-Scale Experiment with a Representative Subject Pool." *Journal of Economic Behavior and Organization* 78 (3): 349–65.
- Bénabou, Roland.** 2008. "Ideology: Joseph Schumpeter Lecture." *Journal of the European Economic Association* 6 (2–3): 321–52.
- Bénabou, Roland, and Jean Tirole.** 2006. "Belief in a Just World and Redistributive Politics." *Quarterly Journal of Economics* 121 (2): 699–746.
- Ben-Ner, Avner, Louis Putterman, Dan Magen, and Fanmin Kong.** 2004. "Reciprocity in a Two-Part Dictator Game." *Journal of Economic Behavior and Organization* 53 (3): 333–52.
- Benoit, Jean-Pierre, and Juan Dubra.** 2011. "Apparent Overconfidence." *Econometrica* 79 (5): 1591–1625.
- Berg, Joyce, John W. Dickhaut, and Kevin A. McCabe.** 1995. "Trust, Reciprocity, and Social History." *Games and Economic Behavior* 10 (1): 122–42.
- Bisin, Alberto, and Thierry Verdier.** 2000. "A Model of Cultural Transmission, Voting and Political Ideology." *European Journal of Political Economy* 16 (1): 5–29.
- Blount, Sally.** 1995. "When Social Outcomes Aren't Fair: The Effect of Causal Attributions on Preferences." *Organizational Behavior and Human Decision Processes* 63 (2): 131–44.
- Bolton, Gary E., Elena Katok, and Rami Zwick.** 1998. "Dictator Game Giving: Rules of Fairness versus Acts of Kindness." *International Journal of Game Theory* 27 (2): 269–99.
- Bolton, Gary E., and Axel Ockenfels.** 2000. "ERC: A Theory of Equity, Reciprocity, and Competition." *American Economic Review* 90 (1): 166–93.
- Bowles, Samuel.** 1998. "Endogenous Preferences: The Cultural Consequences of Markets and Other Economic Institutions." *Journal of Economic Literature* 36 (1): 75–111.
- Brunnermeier, Markus K., and Jonathan A. Parker.** 2005. "Optimal Expectations." *American Economic Review* 95 (4): 1092–1118.
- Caplin, Andrew, and John Leahy.** 2001. "Psychological Expected Utility Theory and Anticipatory Feelings." *Quarterly Journal of Economics* 116 (1): 55–79.
- Cappelen, Alexander W., Astri Drange Hole, Erik Ø. Sorensen, and Bertil Tungodden.** 2007. "The Pluralism of Fairness Ideals: An Experimental Approach." *American Economic Review* 97 (3): 818–27.
- Cherry, Todd L., Peter Frykblom, and Jason F. Shogren.** 2002. "Hardnose the Dictator." *American Economic Review* 92 (4): 1218–21.
- Compte, Olivier, and Andrew Postlewaite.** 2004. "Confidence-Enhanced Performance." *American Economic Review* 94 (5): 1536–57.
- Corneo, Giacomo, and Hans Peter Gruner.** 2002. "Individual Preferences for Political Redistribution." *Journal of Public Economics* 83 (1): 83–107.
- Cruces, Guillermo, Ricardo Perez-Truglia, and Martin Tetaz.** 2013. "Biased Perceptions of Income Distribution and Preferences for Redistribution: Evidence from a Survey Experiment." *Journal of Public Economics* 98: 100–112.
- Dahl, Gordon B., and Michael R. Ransom.** 1999. "Does Where You Stand Depend on Where You Sit? Tithing Donations and Self-Serving Beliefs." *American Economic Review* 89 (4): 703–27.
- Dal Bó, Pedro, Andrew Foster, and Louis Putterman.** 2010. "Institutions and Behavior: Experimental Evidence on the Effects of Democracy." *American Economic Review* 100 (5): 2205–29.
- Dana, J., Daylian M. Cain, and Robyn M. Dawes.** 2006. "What You Don't Know Won't Hurt Me: Costly (But Quiet) Exit in Dictator Games." *Organizational Behavior and Human Decision Processes* 100 (2): 193–201.
- Dana, Jason, Roberto A. Weber, and Jason Xi Kuang.** 2007. "Exploiting Moral Wiggle Room: Experiments Demonstrating an Illusory Preference for Fairness." *Economic Theory* 33 (1): 67–80.
- Della Vigna, Stefano, John A. List, and Ulrike Malmendier.** 2012. "Testing for Altruism and Social Pressure in Charitable Giving." *Quarterly Journal of Economics* 127 (1): 1–56.
- Denzau, Arthur T., and Douglass C. North.** 1994. "Shared Mental Models: Ideologies and Institutions." *Kyklos* 47 (1): 3–31.
- Di Tella, Rafael, Sebastian Galiani, and Ernesto Schargrodsky.** 2007. "The Formation of Beliefs: Evidence from the Allocation of Land Titles to Squatters." *Quarterly Journal of Economics* 122 (1): 209–41.
- Di Tella, Rafael, and Robert MacCulloch.** 2009. "Why Doesn't Capitalism Flow to Poor Countries?" *Brookings Papers on Economic Activity* (Spring): 285–321.
- Di Tella, Rafael, and Ricardo Perez-Truglia.** 2010. "Conveniently Upset: Avoiding Altruism by Distorting Beliefs about Others." National Bureau of Economic Research Working Paper 16645.

- Di Tella, Rafael, Ricardo Perez-Truglia, Andres Babino, and Mariano Sigman.** 2015. "Conveniently Upset: Avoiding Altruism by Distorting Beliefs about Others' Altruism: Dataset." *American Economic Review*. <http://dx.doi.org/10.1257/aer.20141409>.
- Dunning, D., J. A. Meyerowitz, and A. D. Holzberg.** 1989. "Ambiguity and Self-Evaluation: The Role of Idiosyncratic Trait Definitions in Self-Serving Assessments of Ability." *Journal of Personality and Social Psychology* 57 (6): 1082–90.
- Dusek, L., A. Ortmann, and L. Lízal.** 2004. "Understanding Corruption through Experiments: A Primer." CERGE-EI Discussion Paper 136.
- 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 Schmidt, Klaus M.** 1999. "A Theory of Fairness, Competition, and Cooperation." *Quarterly Journal of Economics* 114 (3): 817–68.
- Festinger, Leon.** 1957. *A Theory of Cognitive Dissonance*. Stanford: Stanford University Press.
- Fischbacher, U.** 1999. "z-Tree: Toolbox for Ready-Made Economic Experiments." IEW Working Paper 21.
- Fong, Christina.** 2001. "Social Preferences, Self-Interest, and the Demand for Redistribution." *Journal of Public Economics* 82 (2): 225–46.
- Forsythe, Robert, Joel L. Horowitz, N. E. Savin, and Martin Sefton.** 1994. "Fairness in Simple Bargaining Experiments." *Games and Economic Behavior* 6 (3): 347–69.
- Giuliano, Paola.** 2007. "Living Arrangements in Western Europe: Does Cultural Origin Matter?" *Journal of the European Economic Association* 5 (5): 927–52.
- Giuliano, Paola, and Antonio Spilimbergo.** 2014. "Growing Up in a Recession." *Review of Economic Studies* 81 (2): 787–817.
- Haisley, Emily C., and Roberto A. Weber.** 2010. "Self-Serving Interpretations of Ambiguity in Other-Regarding Behavior." *Games and Economic Behavior* 68 (2): 614–25.
- Hamman, John R., George Loewenstein, and Roberto A. Weber.** 2010. "Self-Interest through Delegation: An Additional Rationale for the Principal-Agent Relationship." *American Economic Review* 100 (4): 1826–46.
- Hastorf, Albert H., and Hadley Cantril.** 1954. "They Saw a Game: A Case Study." *Journal of Abnormal and Social Psychology* 49 (1): 129–34.
- Henrich, Joseph, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, Herbert Gintis, and Richard McElreath.** 2001. "In Search of Homo Economicus: Behavioral Experiments in 15 Small-Scale Societies." *American Economic Review* 91 (2): 73–78.
- Hoffman, Elizabeth, Kevin McCabe, Keith Shachat, and Vernon Smith.** 1994. "Preferences, Property Rights, and Anonymity in Bargaining Games." *Games and Economic Behavior* 7 (3): 346–80.
- Hoffman, Elizabeth, Kevin McCabe, and Vernon L. Smith.** 1996. "Social Distance and Other-Regarding Behavior in Dictator Games." *American Economic Review* 86 (3): 653–60.
- Hoffman, E., and L. Spitzer.** 1985. "Entitlements, Rights and Fairness: An Experimental Examination of Concepts of Distributive Justice." *Journal of Legal Studies* 14 (2): 259–97.
- Hsee, Chris K.** 1996. "Elastic Justification: How Unjustifiable Factors Influence Judgments." *Organizational Behavior and Human Decision Processes* 66 (1): 122–29.
- Kahneman, Daniel, Jack L. Knetsch, and Richard Thaler.** 1986. "Fairness as a Constraint on Profit Seeking: Entitlements in the Market." *American Economic Review* 76 (4): 728–41.
- Konow, James.** 2000. "Fair Shares: Accountability and Cognitive Dissonance in Allocation Decisions." *American Economic Review* 90 (4): 1072–91.
- Konow, James.** 2003. "Which Is the Fairest One of All? A Positive Analysis of Justice Theories." *Journal of Economic Literature* 41 (4): 1188–1239.
- Konow, James.** 2012. "Adam Smith and the Modern Science of Ethics." *Economics and Philosophy* 28 (3): 333–62.
- Köszegi, Botond.** 2006. "Ego Utility, Overconfidence, and Task Choice." *Journal of the European Economic Association* 4 (4): 673–707.
- Kuhn, Andreas.** 2011. "In the Eye of the Beholder: Subjective Inequality Measures and Individuals' Assessment of Market Justice." *European Journal of Political Economy* 27 (4): 625–41.
- Kunda, Ziva.** 1990. "The Case for Motivated Reasoning." *Psychological Bulletin* 108 (3): 480–98.
- 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.
- Lerner, Melvin J.** 1982. *The Belief in a Just World: A Fundamental Delusion*. New York: Springer.
- Levine, David K.** 1998. "Modeling Altruism and Spitefulness in Experiments." *Review of Economic Dynamics* 1 (3): 593–622.
- Malmendier, Ulrike, Vera L. te Velde, and Roberto A. Weber.** 2014. "Rethinking Reciprocity." *Annual Review of Economics* 6: 849–74.

- Mobius, M., M. Niederle, P. Niehaus, and T. Rosenblat.** 2014. "Managing Self-Confidence." Unpublished.
- Norton, M., and D. Ariely.** 2011. "Building a Better America: One Wealth Quintile at a Time." *Perspectives on Psychological Science* 6 (9): 9–12.
- Oberholzer-Gee, Felix, and Reiner Eichenberger.** 2008. "Fairness in Extended Dictator Game Experiments." *B.E. Journal of Economic Analysis and Policy* 8 (1): 1–21.
- Oxoby, Robert J.** 2003. "Attitudes and Allocations: Status, Cognitive Dissonance, and the Manipulation of Attitudes." *Journal of Economic Behavior and Organization* 52 (3): 365–85.
- Oxoby, Robert J., and John Spraggon.** 2008. "Mine and Yours: Property Rights in Dictator Games." *Journal of Economic Behavior and Organization* 65 (3-4): 703–13.
- Piketty, Thomas.** 1995. "Social Mobility and Redistributive Politics." *Quarterly Journal of Economics* 110 (3): 551–84.
- Rabin, Matthew.** 1993. "Incorporating Fairness into Game Theory and Economics." *American Economic Review* 83 (5): 1281–1302.
- Rabin, Matthew.** 1995. "Moral Preferences, Moral Constraints, and Self-Serving Biases." University of California at Berkeley Economics Working Paper 95-241.
- Rabin, Matthew, and Joel L. Schrag.** 1999. "First Impressions Matter: A Model of Confirmatory Bias." *Quarterly Journal of Economics* 114 (1): 37–82.
- Rotemberg, Julio J.** 2005. "Customer Anger at Price Increases, Changes in the Frequency of Price Adjustment and Monetary Policy." *Journal of Monetary Economics* 52 (4): 829–52.
- Rotemberg, Julio J.** 2008. "Minimally Acceptable Altruism and the Ultimatum Game." *Journal of Economic Behavior and Organization* 66 (3-4): 457–76.
- Rotemberg, Julio J.** 2014. "Group Learning, Wage Dispersion, and Nonstationary Offers." Unpublished.
- Roth, Alvin E.** 1995. "Bargaining Experiments." In *The Handbook of Experimental Economics*, edited by John H. Kagel and Alvin E. Roth, 253–348. Princeton: Princeton University Press.
- Ruffle, Bradley J.** 1998. "More Is Better, But Fair Is Fair: Tipping in Dictator and Ultimatum Games." *Games and Economic Behavior* 23: 247–65.
- Schweitzer, Maurice E., and Christopher K. Hsee.** 2002. "Stretching the Truth: Elastic Justification and Motivated Communication of Uncertain Information." *Journal of Risk and Uncertainty* 25 (2): 185–201.
- Sharot, Tali, Christoph Korn, and Raymond Dolan.** 2011. "How Unrealistic Optimism Is Maintained in the Face of Reality." *Nature Neuroscience* 14 (11): 1475–79.