This work is distributed as a Discussion Paper by the

STANFORD INSTITUTE FOR ECONOMIC POLICY RESEARCH

SIEPR Discussion Paper No. 07-30

Social Image and the 50-50 Norm
A Theoretical and Experimental Analysis of Audience Effects

By
James Andreoni
Occidental College
And
B. Douglas Bernheim
Stanford University

August 2007

Stanford Institute for Economic Policy Research
Stanford University
Stanford, CA 94305
(650) 725-1874

Social Image and the 50-50 Norm

A Theoretical and Experimental Analysis of Audience Effects*

James Andreoni, University of California, San Diego
B. Douglas Bernheim, Stanford University and NBER

August 18, 2007

Abstract

A norm of 50-50 division appears to have considerable force in a wide range of economic environments, both in the real world and in the laboratory. Even in settings where one party unilaterally determines the allocation of a prize (the dictator game), many subjects voluntarily cede exactly half to another individual. The hypothesis that people care about fairness does not by itself account for key experimental patterns – for example, that there is frequently a gap in the distribution of transfers just below 50%, or that the frequency of 50-50 splits is sensitive to observability and social distance. We consider an alternative explanation, which adds the hypothesis that people like to be *perceived* as fair. The properties of equilibria for the resulting signaling game correspond closely to laboratory observations. The theory has additional testable implications, and we confirm the validity of these implicaitons through new experiments.

*We would like to thank Iris Bohnet, Colin Camerer, Navin Kartik, and Antonio Rangel for helpful comments, as well as seminar participants at the California Institute of Technology and Stanford University's SITE workshop in Psychology and Economics. We acknowledge financial support from the National Science Foundation through grant numbers SES-0551296 (Andreoni), and SES-0452300 (Bernheim).

1 Introduction

In laboratory experiments, economists usually view audience effects as unfortunate confounds that obscure "real" motives. We often attempt to eliminate these effects, e.g., through the use of double-blind designs. Yet both casual observation and honest introspection strongly suggest that audience effects are pervasive in real economic choices. People care deeply about how others perceive them, and these concerns influence a wide range of decisions. Thorstein Veblen's theory of conspicuous consumption exemplifies these principles, but the phenomenon appears to be much more general.[1] In some instances, concerns for social image arise for instrumental reasons. Yet these concerns remain even when instrumental explanations are absent – for example, most of us care about the impressions we make on complete strangers, even when we don't expect to encounter them again. As a result of either biology or social conditioning, we simply feel good when others think highly of us. Given the likelihood that this effect contributes to many economic phenomenon, it is important to develop and test theories of audience effects, rather than to focus attention on unrealistic settings in which these effects are absent.

In this paper, we examine one particularly widespread phenomenon: the norm of 50-50 division (or, more generally, equal division with more than two parties). This norm appears to have considerable force in a wide range of economic environments. Fifty-fifty sharing rules are well-documented in the context of asymmetric joint ventures among corporations (e.g. Veuglers and Kesteloot [1996], Dasgupta and Tao [1998], and Hauswald and Hege [2003]),[2] share tenancy in agriculture (e.g. De Weaver and Roumasset [2002], Agrawal [2002]), and bequests to children (e.g. Wilhelm [1996], Menchik [1980, 1988]). "Splitting the difference" is a common outcome in negotiations and conventional arbitration (Bloom [1986]). Casual observation suggests other examples: business partners often split the earnings from joint projects equally provided their contributions are roughly similar; friends often split

[1] See Bagwell and Bernheim [1996] for a modern treatment of Veblen's theory.

[2] Where issues of control are critical, one also commonly sees a norm of 50-plus-one-share.

restaurant tabs equally provided all have ordered roughly comparable meals, and the U.S. government splits the nominal burden of the payroll tax equally between employers and employees.

The tendency to comply with a 50-50 norm has been duplicated in the laboratory. Even in the extreme case where one party has all of the bargaining power (the "dictator game"), typically 20 to 30 percent of subjects voluntarily cede half of a fixed payoff to another individual (Camerer [1997]); in "ultimatum games," where the second individual has the opportunity to reject the offer (in which case both subjects receive nothing), the fraction offering a 50-50 split is considerably higher (see e.g. the review of experimental results in Fehr and Schmidt [1999]).

For the dictator game, there is clear evidence that the strength of the 50-50 norm is related to audience effects. Choices are sensitive to observability and social distance. In double-blind trials, the typical subject cedes a small amount, and significantly fewer adhere to the 50-50 norm (e.g. Hoffman et. al. [1996]). Conversely, when dictators and recipients face each other, adherence to the 50-50 norm is far more common (Bohnet and Frey [1999]).[3]

A number of theories have been proposed to account for the prevalence of 50-50 division in dictator games. Some of the leading alternatives invoke altruism or concerns for fairness. Neither approach – by itself – accounts for sensitivity of choice to observability and social distance. Other puzzling features of the experimental data underscore our limited understanding of behavior even in this simple environment. In particular:

- There is frequently a trough in the distribution of fractions ceded just below 50%. For example, in Forsythe et. al. [1994], subjects were given ten dollar bills, and were asked to divide these with another individual; 17% ceded one dollar, 13% ceded two dollars, 29% ceded three dollars, 0% ceded four dollars, and 21% ceded five dollars. To account

[3] Several economic experiments have found that audiences enhance generosity. Andreoni and Petrie [2004] and Rege and Telle [2004] find a greater tendency to equalize payoffs when there is an audience. Studies of field data confirm that an audience increases charitable giving (Soetevent [2005]). Indeed, charities can influence contributions by adjusting the coarseness of the information provided to the audience (Harbaugh's [1998]).

for this gap within the context of standard theories, one would need to assume rather peculiar distributions for underlying preference parameters.

- A significant fraction of the population elects precisely 50-50 division, even when it is possible to give slightly less or slightly more. In the context of the fairness model, this would require a kink in the utility function at equal division (which is indeed what Fehr and Schmidt [1999] assume), or approximate optimization. Both assumptions are somewhat awkward. In the context of the altruism model, the same evidence would require a fortuitous atom in the distribution of the critical preference parameter.

- It is extremely rare for a subject to cede more than 50% of the aggregate payoff. The fairness model accounts for this pattern, but the altruism model again would require a peculiar population distribution for the key preference parameter.

- In a dictator game with a total prize of $\$X$, nearly two-thirds of dictators are willing to exit the game for a payoff of $\$Y$ with $Y < X$, provided that the recipient is not told about the game (Brobert et. al. [2007]; see also Dana et. al. [2006]). Neither the altruism model nor the fairness model account for these inefficient choices.

The hypothesis that people care about fairness is plausible, and probably helps to account for behavior both in simple laboratory settings and in the real world. However, it does not appear to be enough. This paper explores the implications of supplementing this hypothesis with an additional, plausible assumption: people like to be perceived as fair. In the model described below, we incorporate the desire to be perceived as fair into the utility function; however, one could also model this dependence as arising explicitly from concerns about subsequent interactions. Our theory explains (1) why 50-50 emerges as a norm, (2) why the force of the norm is related to observability and social distance, (3) why there is a gap in the distribution of choices below 50%, (4) why people choose *exactly* 50-50, even when it is possible to give slightly more or slightly less, (5) why very few people give more than 50%, and (6) why a dictator would sacrifice part of the total prize to opt out of the game.

The intuition for each of these results is straightforward. Our model gives rise to a signaling game wherein the dictator's choice affects others' inferences about his tastes for fairness, and where he likes others to think he's fair. For standard reasons, in any separating equilibrium the dictator tries to create a favorable impression by transferring more than he would if he were not concerned about others' inferences. This is sustainable because the cost of giving more away is lower for those who attach more importance to fairness. However, there's a catch: the single crossing property is not satisfied in this setting. Once the transfer exceeds half of the total payoff, giving more away *increases* inequality, and is therefore *more* costly on the margin for those who attach more importance to fairness. Dictators try to distinguish themselves from those who care less about fairness by giving more, but this only works up to the point where the transfer equals half of the total pie; further separation is impossible. Accordingly, one obtains an atom in the population distribution of choices at a 50-50 split, even when there are no atoms in the population distribution of preferences (explaining items (1) and (4)), and no dictator gives away more than half the prize (explaining item (5)).[4] As is typical in signaling models with partial pooling, there is a gap in the distribution of choices immediately below any atom (explaining item (2)); otherwise, there would be a discontinuity in inferences (and hence perceived fairness) right below the action associated with the atom, and this would induce those choosing slightly lower actions to deviate. A downward shift in the distribution of weights attached to perceptions of fairness (e.g. because opportunities for reciprocation are fewer) moves individuals toward the separating region and away from the pooling region, reducing both the average transfer and the measure of types pooling at 50-50 (explaining item (3)). Finally, many dictators will prefer to act selfishly as long as they can do so without triggering negative inferences (explaining item (6)).

It is important to emphasize that this is not necessarily the best explanation – or even a good explanation – for all 50-50 norms observed in practice, or even all those mentioned

[4] Similar properties would hold if we simply imposed 50% as a ceiling on transfers (for the reasons explored in Cho and Sobel [1990]); here, the ceiling is endogenous.

at the outset of this section.[5] Nevertheless, it deserves serious consideration in many cases. With respect to joint ventures, for example, corporate managers may accede to a 50-50 rule even when they are in relatively strong bargaining positions because they believe this will help establish an atmosphere of trust wherein their partner will expect them to act fairly in the future, and where this expectation will induce the partner to invest more effort and resources in the venture.

We also derive theoretical results for an alternative version of the dictator game that lead to additional testable implications. The alternative setting works as follows. At the outset, nature chooses $x = x_0$ (some particular value) with probability p, in which case the game is over, or selects the dictator game with probability $1 - p$. While the dictator observes nature's choice, an audience (including but not necessarily limited to the responder) does not; the audience does not know whether the outcome is the result of nature's choice or the dictator's choice. For the case of $x_0 = 0$, we show that an increase in p strictly increases the mass of dictators selecting zero, and strictly reduces (if positive) the mass of dictators selecting equal division. Similar properties hold as long as x_0 is close to zero. Accordingly, by varying the parameters of this extended game (p and x_0), we should be able to create a spike in the distribution of voluntary choices at or near zero, and manipulate both its size and location. In contrast, if social image is not a concern, varying these parameters should have no effect on behavior.

[5] Bernheim and Severinov [2003] propose an explanation for equal division of bequests that is also based on partial pooling in a signaling model. The mechanism appropriate for bequests is quite different from that considered here, and is probably not applicable to many other contexts in which 50-50 norms are observed (likewise, the mechanism considered here is probably not applicable to bequests). In their model, parents differ according to their relative affection for two children, and children care about parental affection, which they infer from actions. The direction of imitation changes as one moves from one end of the type space to another, which is why parent types pile up on some intermediate action. This mechanism is not a plausible explanation for what transpires in the dictator game; to invoke it, one would need to assume that those who care little about the receiver want to be perceived as caring more, and that those who care a lot about the receiver want to be perceived as caring less. In ultimatum and dictator games, the direction of imitation does *not* change as one moves from one end of the type space to the other – everyone wishes to be perceived as more fair. The main result here is also cleaner in the following sense: in Bernheim and Severinov [2003], there is a range of possible equilibrium norms that includes equal division; here (as we'll see), equal division is the only possible equilibrium norm.

We examine the validity of these testable implications by conducting new experiments. Subjects manifestly exhibit the predicted behavior, and to a striking degree. This corroborates our particular theory, and more generally supports a signaling interpretation of audience effects.

Broadly, this paper falls within the literature on "psychological games," in which players are assumed to have preferences over the beliefs of others (as in Geanakoplos, Pearce, and Stecchetti [1989]). More narrowly, it is related to other work that explores the behavioral implications of concern for social image (e.g., Bernheim [1994], Bagwell and Bernheim [1996]). Recent papers in this general area include work on esteem (Ellingsen and Johannesson [2006]), shame (Tadelis [2007]), and respect (Manning [2007]).

The remainder of this paper is organized as follows. Section 2 describes the model. Section 3 examines some preliminary issues (including the failure of single crossing). Section 4 studies equilibria of the basic model. Section 5 considers testable comparative statics. Section 6 describes our experimental procedures and results. Section 8 offers some conclusions. Proofs of theorems appear in the appendix.

2 The Model

There are two players, a dictator (D) and a receiver (R). D controls a prize normalized to have unit value, and divides it between himself and R. Let $x \in [0, 1]$ denote the amount given to R. D consumes $c = 1 - x$. D is drawn from a population of potential players, differentiated by the parameter t, which indicates the importance placed on fairness. Potential types t lie in the set $[0, \bar{t}]$, and the distribution is given by some atomless CDF H.[6] We define H_s as the CDF obtained from H, conditioning on $t \geq s$. The value of t is D's

[6] Some dictator game experiments appear to produce an atom in the choice distribution at 0, though the evidence for this pattern is mixed (see e.g., Camerer [2003]). Our model does not produce this pattern unless we assume that there is an atom in the distribution of types at $t = 0$. Since the type space is truncated below at 0, it may be reasonable to allows for this possibility. One could also generate a choice atom at zero by assuming that some individuals do not care about social image (in whice case the analysis would be more similar to that presented in Section 6). In experiments, it is also possible that a choice atom at zero results from the discreteness of the choice set and/or approximate optimization.

private information. D cares about his own prize (c), the fairness of the distribution (g), and perceptions of his fairness (m) by some audience (A), which includes (but is not necessarily limited to) R:[7]

$$U(c, m, g, t) = F(c, m) + tg$$

The function F is taken to be unbounded (in both arguments), twice continuously differentiable, strictly increasing (with, for some $f > 0$, $F_1(c, m) > f$ for all $c \in [0, 1]$ and $m \in \mathbb{R}_+$), and strictly concave in c. Thus, payoff increases in own consumption, D's perceived fairness, and fairness of the outcome.

Social image m depends on A's perception of D's fairness. People who are more fair are accorded more respect. If A is certain that D's type is $\widehat{t}$, we will assume that $m = \widehat{t}$.[8] We need to allow for cases where A is uncertain about D's type. Let Φ denote the CDF representing A's beliefs about D's type. We suppose that $m = B(\Phi)$ for some function B satisfying the following weak assumption:

Assumption B-1: (1) B is continuous (where the set of CDFs is endowed with the weak topology). (2) min $supp(\Phi) \leq B(\Phi) \leq$ max $supp(\Phi)$, with strict inequalities when the support of Φ is non-degenerate. (Thus, if Φ places probability one on type $\widehat{t}$, then $B(\Phi) = \widehat{t}$). (3) If Φ' is "higher" than Φ'' in the sense of first-order stochastic dominance, then $B(\Phi') > B(\Phi'')$.

As an example, B might calculate the mean of t given Φ. In effect, assumption B-1 requires that B is a well-behaved aggregator.

Though the players are asymmetric in terms of bargaining power, we assume that they are symmetric with respect to publicly observed indicia of merit. Accordingly, the fairness of an outcome is taken to depend on the extent to which it departs from equal division.[9] We

[7] In experimental settings, the audience may include the experimenter and other subjects.

[8] This is just a normalization; m could be an increasing and continuous function of $\widehat{t}$, in which case that function has been incorporated into F.

[9] If the players were asymmetric with respect to publicly observed indicia of merit, the fairness of an outcome might depend on the extent to which it departed from some other benchmark, such as 60-40. Similar results would follow, except that the behavioral norm would correspond to this alternate benchmark, rather to 50-50.

assume that $g = G(x - \frac{1}{2})$, where the function G is twice continuously differentiable, strictly concave, and reaches a maximum at zero. Thus, we can rewrite D's payoff as

$$U(x, m, t) = F(1 - x, m) + tG\left(x - \frac{1}{2}\right)$$

Play is simple. D selects x. Seeing x, R forms an inference Φ about t. D does not see this inference directly, but knows the equilibrium relation between x and Φ, and therefore accounts for the effect of the decision on his perceived fairness. This game involves a form of signaling, but, as we'll see, the single-crossing property is not satisfied, so the solution is non-standard.

A signaling equilibrium consists of a mapping Q from types (t) to decisions (x), and a mapping P from decisions (x) to inferences (Φ), where decisions are optimal given inferences, and inferences are consistent with decisions. We will use Φ_x to denote the inference associated with the action x. We will confine our attention to pure strategy equilibria.

3 Preliminary Observations

3.1 Solution when type is observable

Suppose for the moment that type is observable. What will D choose? Let $x^*(t)$ denote the first-best choice. Since G is strictly concave and F is strictly concave in c, the best choice is unique, and a continuous function of the preference parameters.

The following theorem tells us what happens if type is observable. No one chooses $x = \frac{1}{2}$. If image and prize are weak substitutes, the size of the gift is weakly increasing in t. There is always mass at zero. Finally, the optimal choice converges to $\frac{1}{2}$ as t gets large. See Figure 1 for an illustration.

Theorem 1: (1) For all t, $x^*(t) \in [0, \frac{1}{2})$. (2) If $F_{12} \leq 0$, then $x^*(t)$ is weakly increasing in t (strictly when $x^*(t) \in (0, 1)$). (3) There exists $t^* > 0$ such that $x^*(t) = 0$ for $t \leq t^*$. (4) $\lim_{t \to \infty} x^*(t) = \frac{1}{2}$.

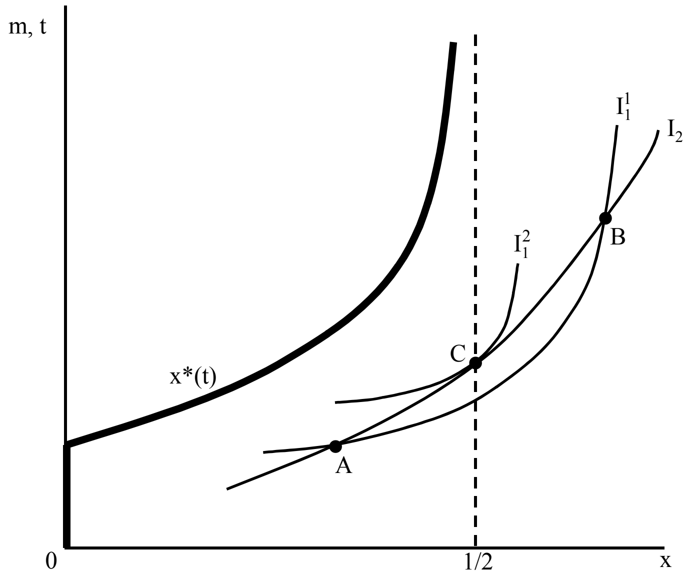

Figure 1: Optimal Choices with Complete Information, and Indifference Curves

Example Let $F(1-x, m) = 1 - x + \alpha m$, and let $G\left(x - \frac{1}{2}\right) = -\left(x - \frac{1}{2}\right)^2$. Then $x^*(t) = \max\left\{0, \frac{1}{2} - \frac{1}{2t}\right\}$. Note that $x^*(t) = 0$ for $t < 1$.

Even though we've assumed $F_2 > 0$, the proof of Theorem 1 subsumes the case where social image, m, doesn't matter ($F_2 = 0$). So, as claimed in Section 1, a model with concern for fairness, but without social image, produces a counterfactual implication – it predicts that everyone should select a transfer strictly less than $\frac{1}{2}$.

3.2 The failure of single-crossing

For the usual reasons, the solution described in Theorem 1 is not sustainable as a separating equilibrium when type is inferred from action, and when social image depends on the inferred type. Here we lay the groundwork for an analysis of signaling equilibria.

To begin, we examine D's willingness to make trade-offs between the magnitude of the gift and social image. Setting $U(x, m, t)$ equal to a constant and differentiating implicitly, we obtain

$$\left.\frac{dm}{dx}\right|_U = -\frac{tG'\left(x-\frac{1}{2}\right) - F_1\left(1-x,m\right)}{F_2\left(1-x,m\right)}$$

Suppose that $x < \frac{1}{2}$. Then $G'\left(x-\frac{1}{2}\right) > 0$. An increase in t causes $\left.\frac{dm}{dx}\right|_U$ to fall. In other words, in the (x,m) plane, indifference curves, if positively sloped, become flatter with an increase in t, and, if negatively sloped, become steeper. This is the usual Spence-Mirrlees condition. Increasing x is, in effect, less costly to those with high values of t, so these types can signal their fairness by giving more.

However, for $x > \frac{1}{2}$, $G'\left(x-\frac{1}{2}\right) < 0$. An increase in t causes $\left.\frac{dm}{dx}\right|_U$ to rise. In other words, in the (x,m) plane, indifference curves, if positively sloped, become steeper with an increase in t, and, if negatively sloped, become flatter. This is the usual Spence-Mirrlees condition in reverse. Increasing x is, in effect, more costly to those with high values of t, so these types cannot signal their fairness by giving more.

We illustrate these points in Figure 1. The indifference curves I_1^1 and I_1^2 belong to type t_1, while the indifference curve I_2 belongs to type t_2, where $t_1 > t_2$. Notice that I_1^1 is flatter than I_2 at A, where $x < \frac{1}{2}$, and steeper than I_2 at B, where $x > \frac{1}{2}$. Also, I_1^2 is tangent to I_2 at C, where $x = \frac{1}{2}$.

Given these properties, one can see that, intuitively, in a signaling equilibrium, $x = \frac{1}{2}$ naturally serves as something of a barrier. Of course, it is not literally a barrier, and indeed there are equilibria where some type or types choose to transfer more than $\frac{1}{2}$. However, there is only limited scope in equilibrium for choices exceeding $\frac{1}{2}$ (see Lemma 2 in the appendix), and these possibilities do not survive the application of standard refinements.

4 Equilibrium

As in most signaling models, there are many equilibria. In this section, we describe a particular class of equilibria. We show that existence is guaranteed and demonstrate uniqueness within this class. We also show that it is reasonable to focus on the identified equilibrium because it is the only one satisfying the D1 criterion (a standard equilibrium refinement for signaling games). We also discuss some of the properties of this equilibrium.

4.1 Separation

Since the problem has a fairly standard signaling structure for $x \in \left[0, \frac{1}{2}\right]$, it's natural to start by looking for separating equilibria.

In a separating equilibrium with action function $S(t)$, action x leads to inference $\varphi(x) = S^{-1}(x)$ (provided this action is taken by some type). Optimality requires each type t to choose a values of x that maximizes the function $U(x, S^{-1}(x), t)$. Taking the first order condition, and then substituting $x = S(t)$ (which holds in any separating equilibrium), we have:

$$S'(t) = -\frac{F_2\left(1 - S(t), t\right)}{tG'\left(S(t) - \frac{1}{2}\right) - F_1\left(1 - S(t), t\right)} \tag{1}$$

The preceding is a non-linear first order differential equation. As usual, given an initial condition (a choice z and an associated type r), a unique solution to this equation is guaranteed. As we will see below, in all cases of interest the solution is strictly increasing, so the inverse function (the inference function) is also well defined.

We will denote the solution with initial condition (r, z) (that is, type r chooses action z) as $S_{r,z}(t)$, and the inverse as $\varphi_{r,z}(x)$ (so that $\varphi_{r,z}(z) = r$ and $S_{r,z}(r) = z$). For the most part, we will find it more useful to discuss the separating action function, rather than the inference function. In general, we will be interested in cases where $z \geq x^*(r)$. This holds, for example, when the process is initialized by letting type $t = 0$ choose $x = 0$.

The following result tells us that, for $z \geq x^*(r)$, $S_{r,z}(t)$ assigns to each type a transfer exceeding the first-best level, that it is strictly increasing in t, and that it assures mutual non-imitation among all actions not exceeding $\frac{1}{2}$. It also tells us that there is some type to which $S_{r,z}(t)$ assigns equal division, and that $S_{r,z}(t)$ is increasing in z and continuous in r and z.

Theorem 2: Assume $z \geq x^*(r)$. (1) $S_{r,z}(t) > x^*(t)$ for $t > r$. (2) For all $t \geq r$, $S'_{r,z}(t) > 0$. (3) If $S_{r,z}(t') \leq \frac{1}{2}$ and $S_{r,z}(t'') \leq \frac{1}{2}$, type $t' \geq 0$ prefers $(x, m) = (S_{r,z}(t'), t')$ to

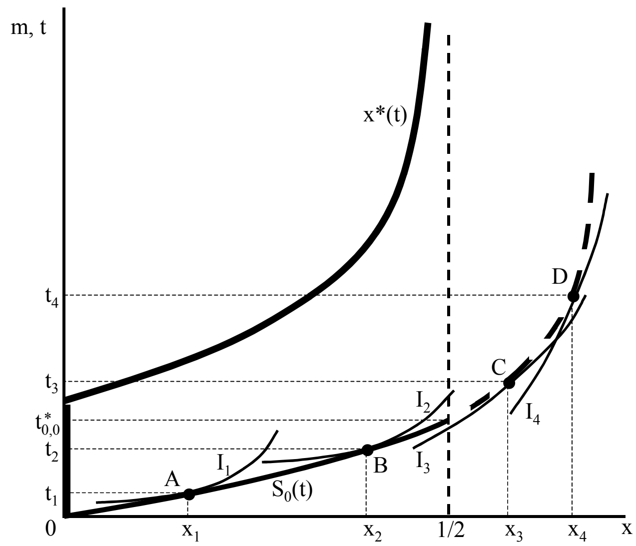

Figure 2: The Separating Function

$(S_{r,z}(t''), t'')$. (4) There exists $t^*_{r,z} > r$ such that $S_{r,z}(t^*) = \frac{1}{2}$. (5) $S_{r,z}(t)$ is increasing in z and continuous in r and z.

We will use $S_0(t)$ as shorthand for the efficient separating function, $S_{0,0}(t)$. We illustrate $S_0(t)$ and $t^*_{0,0}$ (defined by the intersection between $S_0(t)$ and the vertical line at $x = \frac{1}{2}$) in Figure 2. Types t_1 and t_2 choose, respectively, x_1 and x_2. Their indifference curves, I_1 and I_2, are tangent to the curve representing $S_0(t)$ at, respectively, points A and B, so they prefer these alternatives to all other points on the signaling function. Notice that the solution to the differential equation continues to the right of the vertical line at $x = \frac{1}{2}$ (where it is shown as a broken line). For types t_3 and t_4, the solution assigns, respectively, the choices x_3 (point C) and x_4 (point D). Their indifference curves (I_3 and I_4) are tangent to the broken curve at these points, but are on the "wrong side" of it, so the mutual non-imitation constraints are not satisfied. Indeed, the assigned choice are local minima along the broken line, rather than local maxima. Accordingly, the solution only works as a separating function to the left of the vertical line at $x = \frac{1}{2}$.

Now we establish a simple necessary and sufficient condition for the existence of a separating equilibrium.

Theorem 3: A separating equilibrium exists if and only if $\bar{t} \leq t^*_{0,0}$.

Thus, if the population contains people who are "sufficiently" fair-minded, perfect separation is impossible.

4.2 Pooling at equal division

What happens when $\bar{t} > t^*_{0,0}$? Intuitively, if $\frac{1}{2}$ acts as a boundary, we'd expect to find a pool at $\frac{1}{2}$ (Cho and Sobel [1990]). What would such an equilibrium entail? There are two possibilities.

For the first possibility, we divide the types into two segments, $[0, t_0]$ and $(t_0, \bar{t}]$ (with t_0 s.t. $S_0(t_0) < \frac{1}{2}$), and construct the equilibrium as follows: for $t \in [0, t_0)$, $Q(t) = S_0(t)$; for $t \in (t_0, \bar{t}]$, $Q(t) = \frac{1}{2}$. In other words, types separate up to t_0, and all higher types divide the prize equally.

For this configuration to be an equilibrium, the lowest type in the equal-division pool must be indifferent between separating and joining the pool:

$$U(S_0(t_0), t_0, t_0) = U\left(\frac{1}{2}, B\left(H_{t_0}\right), t_0\right)$$

Remember that H_{t_0} is defined as the CDF obtained starting from H (the population distribution) and conditioning on $t \geq t_0$. Because of t_0's indifference, there is a completely identical equilibrium (differing from this one only on a set of measure zero) where t_0 resolves it's indifference in favor of $\frac{1}{2}$ (that is, it joins the pool). Throughout this paper, we adopt the convention of always resolving indifferent in favor of the lower action. This shortens some of the arguments without altering any of the substance (since it always involves a choice among essentially identical equilibria).

We illustrate this type of equilibrium in Figure 3. Notice that the indifference curve for type t_0, labelled I_{t_0}, passes through both point A, the separating choice for t_0, and point B,

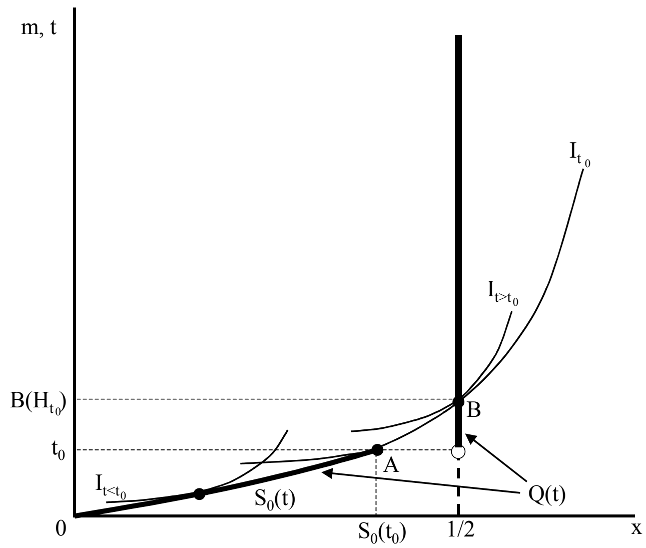

Figure 3: A Central Pooling Equilibrium

the outcome for the pool. For any type $t > t_0$, an indifference curve through point B has the shape of the curve labelled $I_{t>t_0}$; it is flatter than I_{t_0} to the left of B, and steeper to the right. Consequently, all such types strictly prefer the pool to any point on $S_0(t)$ below t_0.

The second possibility is that everyone joins the pool ($Q(t) = \frac{1}{2}$ for all t). This would require

$$U(0,0,0) < U\left(\frac{1}{2}, B\left(H\right), 0\right),$$

so that the lowest type prefers to be in the pool rather than choose his first-best action and receive the worst possible inference.

We refer to these configurations as "central pooling equilibria." The following theorem shows that a central pooling equilibrium exists precisely when a separating equilibrium does not exist. Moreover, when it exists, it is unique. Finally, there is always a gap in the distribution of actions right below $\frac{1}{2}$; in Figure 3, it is the interval $(S_0(t_0), \frac{1}{2})$.

Theorem 4: A non-degenerate central pooling equilibrium exists iff $\overline{t} > t^*_{0,0}$. When this condition holds, the central pooling equilibrium is unique, and $S_0(t_0) < \frac{1}{2}$.

Remark: For $\bar{t} = t^*_{0,0}$, there is a degenerate central pooling equilibrium (the pool consists only of type $\bar{t}$), which is also a separating equilibrium.

Despite some surface similarities, the mechanism producing a central pool in this model differs from those explored in Bernheim [1994] and Bernheim and Severinov [2003]. In those papers, the direction of imitation reverses when type passes some threshold; types in the middle are unable to adjust their choices to simultaneously deter imitation from the left and from the right. Here, higher types always try to deter imitation by lower types, but are simply unable to do this once x reaches $\frac{1}{2}$.

4.3 Justifying the Equilibrium

So far, we have looked at only two types of equilibria: separating equilibria, and equilibria with central pools. We now justify this focus by applying a standard equilibrium refinement: the D1 criterion of Cho and Kreps [1987]. This criterion insists that agents attribute any action not chosen to the type that would be willing to choose it for the widest range of inferences. The following theorem tells us that the separating and central pooling equilibria described above all satisfy the D1 criteria, and indeed are the only equilibria that satisfy this criterion. A similar result holds for other standard criteria (e.g. divinity).

Theorem 5: When $\bar{t} \leq t^*_{0,0}$, the efficient separating equilibrium $S_0(t)$ is the unique signaling equilibrium satisfying the D1 criterion. When $\bar{t} > t^*_{0,0}$, the central pooling equilibrium is the unique signaling equilibrium satisfying the D1 criterion.

In standard signaling environments (with single crossing), the D1 criterion isolates separating equilibria, and equilibria with pools at the upper boundary of the action set (Cho and Sobel [1990]). Consequently, the unusual aspect of this theorem is that central pooling equilibria satisfy the D1 criterion. To understand why, refer back to Figure 3. What inference would the audience make upon observing some $x \in (S_0(t_0), \frac{1}{2})$, or $x > \frac{1}{2}$? The case of $x \in (S_0(t_0), \frac{1}{2})$ is standard. Since the single crossing property is satisfied for $x < \frac{1}{2}$, the indifference curve for t_0, I_{t_0}, lies below the indifference curve of all other types – both $I_{t>t_0}$, and $I_{t<t_0}$. Consequently, t_0 would be willing to choose the action for the widest range of inferences, and therefore, by the D1 criterion, the action would be attributed to t_0. Given this inference, no type would prefer to make this choice. The case of $x > \frac{1}{2}$ is nonstandard. Since the indifference curves of higher types are *steeper* than those of lower types for $x > \frac{1}{2}$, for any such x, the indifference curve of type t_0, I_{t_0} once again lies *below* the indifference curves of all higher types, $I_{t>t_0}$. Consequently, any such choice would be attributed to t_0, or to a lower type. Given this inference, no type would prefer to make this choice.

4.4 Varying the importance of social image

In the laboratory, treatments that make choices observable to larger audiences, or that reduce the social distance between dictators and recipients, generate higher frequencies of 50-50 division. How does our theory account for these patterns?

Presumably, these types of treatments cause dictators to attach greater importance to social image. Formally, we say that $\widetilde{U}$ *attaches more importance to social image* than U if

$$\widetilde{U}(c, m, g, t) = U(c, m, g, t) + \phi(m),$$

where ϕ us differentiable, and $\phi'(m)$ is strictly positive and bounded away from zero. The addition of the separable term $\phi(m)$ allows us to increase the weight attached to social image without altering the trade-off between consumption and equity. We note that the modified model continues to fall without our framework; simply take $\widetilde{F}(c, m) = F(c, m) + \phi(m)$.

The following result tells us that an increase in the importance attached to social image increases the extent to which dictators conform to the 50-50 norm:

Theorem 6: Suppose that $\widetilde{U}$ attaches more importance to social image than U. Let $\widetilde{t}_0$ and t_0 denote the types defining the lower boundary of the central pool for $\widetilde{U}$ and U, respectively. Then $\widetilde{t}_0 \leq t_0$, with strict inequality when $t_0 \in (0, \overline{t})$. Thus, the measure of types choosing $x = \frac{1}{2}$ is weakly greater for $\widetilde{U}$ than for U, and strictly greater when some but not all types choose $x = \frac{1}{2}$ with U.

Theorem 6 confirms that, according to our theory, treatments that make choices observable to larger audiences, or that reduce the social distance between dictators and receivers, should produce greater conformance with the 50-50 norm.

4.5 Observable properties of the equilibrium

Here we provide a brief review of the observed behavioral patterns for which our theory accounts. First, provided that the variance in dictators' preferences are sufficiently great, the model produces a spike in the distribution of choices *precisely* at equal division, even if the prize is perfectly divisible. Second, no one gives more than half of the prize. Third, there is always a gap in the distribution of choices just below equal division. Intuitively, if a dictator intends to divide the pie unequally, it makes no sense to divide it only *slightly* unequally, since the tiny consumption gain will be overwhelmed by negative inferences about his motives. Fourth, conditions that raise the importance attached to social image should result in greater conformance with the 50-50 norm. Thus, the theory accounts for the primary features of the experimental data on the standard dictator game.

5 Further Testable Implications

A good theory not only explains existing observations, but also generates new testable implications. In this section, we explore some sharp new comparative static implications for a modified version of the dictator game. The rest of the paper tests those implications experimentally.

The modified version of the dictator game operates as follows. At the outset, nature chooses some x_0 with probability p, in which case the game is over, or selects the dictator game described in Section 2 with probability $1-p$. While D observes nature's choice, R does not; R does not know whether the outcome is the result of nature's choice or D's choice. As we show in this section, our theory has distinctive implications concerning the effects of x_0 and p on the dictators' choices.

We provide formal results for the case of $x_0 = 0$. With $p > 0$, the distribution of *voluntary* choices has mass at $x = 0$, as well as (in some cases) $x = \frac{1}{2}$. Intuitively, the potential for nature to choose $x = 0$, regardless of the dictator's type, reduces the stigma associated with voluntarily choosing $x = 0$. Moreover, as p increases, more and more dictator types are tempted to "hide" their selfishness behind nature's choice. This mitigates the threat of imitation, thereby allowing higher types to reduce their gifts as well. Accordingly, the measure of types voluntarily choosing $x = 0$ grows, while the measure of types choosing $x = \frac{1}{2}$ shrinks.

Due to space limitations, we do not provide formal results for the case of $x_0 > 0$. However, for any x_0 close to zero, the analysis and intuition are similar.[10] The potential for nature to choose x_0 regardless of the dictator's type reduces the stigma associated with voluntarily choosing x_0. Notably, x_0 becomes more attractive both to those who would otherwise have given more than x_0, and to those who otherwise would have given less. Thus, a pool forms at x_0, rather than at zero, reducing the total mass both above and below x_0. Just as there is a gap in the distribution of voluntary choices just below $x = \frac{1}{2}$, there is also a gap in the distribution of voluntary choices just below x_0 (and for the same reasons).

Accordingly, our theory has two critical comparative static implications. The first concerns p. Setting $p > 0$ creates a second pool (in *voluntary* choices) at x_0; raising p enlarges this pool and shrinks the one at $x = \frac{1}{2}$. There are gaps in the distribution of voluntary choices just below both pools. The second implication concerns x_0: by varying x_0 (near zero), we should be able to manipulate the location of the second pool. Those predictions are specific and testable.

As in Section 4, we proceed by describing a particular class of equilibrium, proving existence and uniqueness within this class, and then showing that the D1 criterion always

[10] When x_0 is close to zero, the equilibrium inference associated with x_0 for $p = 0$ must be close to zero, and therefore less than $B(H)$. Accordingly, setting $p > 0$ (with no change in behavior) leads to a more positive inference for x_0. When x_0 is not close to zero, the equilibrium inference associated with x_0 for $p = 0$ may be greater than $B(H)$, in which case setting $p > 0$ (with no change in behavior) leads to a more negative inference for x_0; hence the same results do not follow.

selects this equilibrium. We then examine comparative statics in p.

5.1 Preliminaries

First we introduce an additional assumption concerning B.

Assumption B-2: Consider the CDFs J, K, and L, such that $J(t) = \lambda K(t) + (1 - \lambda)L(t)$. If max $supp(L) \leq B(K)$, then $B(J) \leq B(K)$, where the second inequality is strict if the first is strict or if the support of L is nondegenerate.

In words, the assumption says the following. If we starts with a distribution K and mix it with a distribution L that places all probability on types whose social images are no higher than the image associated with K, then the social image of the resulting group can't exceed the social image of K. If L also places some probability on types whose social images are lower than the image associated with K, then the social image of the resulting group is lower than the social image of K.

Now we describe a particular class of equilibria. We divide the types into three segments, $[0, t_0]$, $(t_0, t_1]$, and $(t_1, \bar{t}]$, where $t_0 \leq t_1$. We construct equilibria as follows: for $t \in [0, t_0]$, $Q(t) = 0$; for $t \in (t_0, t_1]$, $Q(t) = S_{t_0, x^*(t_0)}(t) \equiv S^{t_0}(t)$ (where the latter definition is to simplify notation); for $t \in (t_1, \bar{t}]$, $Q(t) = \frac{1}{2}$. In other words, types up to t_0 give no gift, types between t_0 and t_1 separate (with t_0 picking his favorite action), and all higher types divide the prize equally.

In this definition, we assign t_0 to the lower pool and t_1 to the middle pool. One could also assign t_0 to the middle pool and/or t_1 to the upper pool. Since each type is of measure zero, we regard these equilibria as equivalent, and always resolve indifference toward the lower choice as a matter of convention.

There are two main cases to consider, and some subcases.

Case A: $t_0 < t_1$ (so the middle segment is present). For this case, there are two subcases.

Case A-1: $t_1 < \bar{t}$ (so the top segment is also present).

Case A-2: $t_1 = \bar{t}$ (so the top segment is not present).

Case B: $t_1 = t_0$ (so the middle segment is not present).

What do we require for these configurations to be equilibria? For Case A (both subcases), type t_0 must be indifferent between joining the pool choosing zero, and separating (which, it turns out, must involve t_0's first-best choice, $x^*(t_0)$):

$$U\left(0, B\left(\widehat{H}_{t_0}\right), t_0\right) = U\left(x^*(t_0), t_0, t_0\right) \tag{2}$$

where $\widehat{H}_{t_0}$ is the CDF for types in the pool choosing zero (H with probability p, and H truncated above at t_0 with probability $1-p$).[11] For Case A-1, type t_1 must be indifferent between joining the pool choosing $\frac{1}{2}$ and separating:

$$U\left(\frac{1}{2}, B\left(H_{t_1}\right), t_1\right) = U\left(S^{t_0}(t_1), t_1, t_1\right) \tag{3}$$

For Case B, type t_0 must be indifferent between pooling at zero and pooling at $\frac{1}{2}$, and (for reasons that will become clear below) must weakly prefer both to its optimal choice with revelation of its type:

$$U\left(0, B\left(\widehat{H}_{t_0}\right), t_0\right) = U\left(\frac{1}{2}, B\left(H_{t_0}\right), t_0\right) \geq U\left(x^*(t_0), t_0, t_0\right) \tag{4}$$

We refer to these configurations as "double pooling equilibria" (even though Case A-2 involves a single pool). Figure 4 illustrates Case A-1 (with two pools and a region of separation). The indifference curve I_{t_0} indicates that type t_0 is indifferent between the lower pool (point A) and separating with it's first best choice, $x^*(t_0)$ (point B). All types between t_0 and t_1 choose a point on the separating function generated when point B is used as the initial condition. The indifference curve I_{t_1} indicates that type t_1 is indifferent between separating (point C) and the upper pool at $x = \frac{1}{2}$ (point D).

5.2 Formal results

Our first result establishes the existence and uniqueness of a double pooling equilibrium. It also characterizes the conditions under which Cases A-1, A-2, and B prevail.

[11] In the proof of Theorem 7, we provide an explicit formula for $\widehat{H}_{t_0}$.

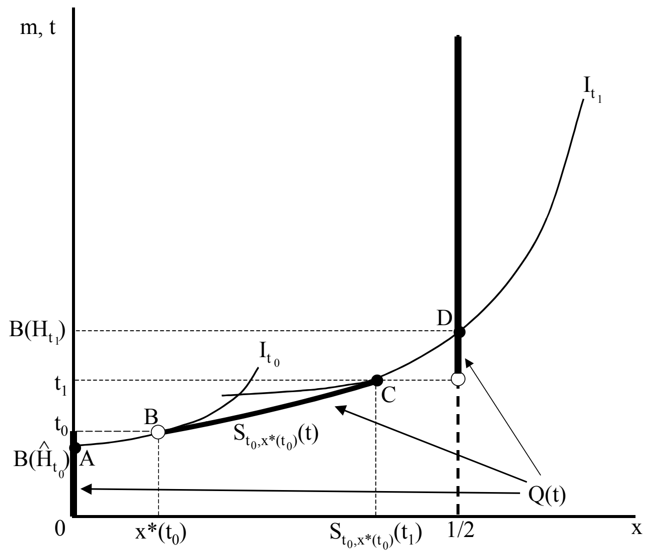

Figure 4: A Double Pooling Equilibrium

Theorem 7: There exists a unique double pooling equilibrium, which is characterized as follows. Equation (2) has a unique solution, $t_0^* \in (0, \bar{t})$. If

$$U\left(x^*(t_0^*), t_0^*, t_0^*\right) \leq U\left(\frac{1}{2}, B\left(H_{t_0^*}\right), t_0^*\right), \tag{5}$$

then the equilibrium belongs to Case B, and $t_0 \in (0, t_0^*)$ (so that there is positive mass at gifts of zero and $\frac{1}{2}$). If

$$U\left(x^*(t_0^*), t_0^*, t_0^*\right) > U\left(\frac{1}{2}, B\left(H_{t_0^*}\right), t_0^*\right), \tag{6}$$

then the equilibrium belongs to Case A, and $t_0 = t_0^* \in (0, \bar{t})$ (so that there is positive mass at zero gift, but not full mass). If in addition $S^{t_0^*}(\bar{t}) > \frac{1}{2}$, it is a Case A-1 equilibrium, and $S^{t_0^*}(t_1) < \frac{1}{2}$ (so that there is positive mass at $\frac{1}{2}$ and a gap in the action distribution just below $\frac{1}{2}$). If $S^{t_0^*}(\bar{t}) \leq \frac{1}{2}$, it is a case A-2 equilibrium.

As in Section 4, we justify our focus on the equilibria described in Theorem 7 by applying the D1 refinement.

Theorem 8: The double-pooling equilibrium is the unique signaling equilibrium satisfying the D1 criterion.

The unique D1 equilibrium of this model has a number of notable properties. For *voluntary* choices, there is always mass at $x = 0$. Nature's exogenous choice of $x = 0$ induces players to "hide" their selfishness by mimicking this choice. There is never positive mass at any other choice except $\frac{1}{2}$. As before, there is a gap in the distribution of choices just below $\frac{1}{2}$. As shown in Figure 4, there may also be a gap in the distribution of choices just above $x = 0$, but this gap only forms for large values of p.[12]

Now we turn to the role of p. It turns out that both t_0 and t_1 are monotonically increasing in p. This means that the mass at $x = 0$ grows, and the mass at $x = \frac{1}{2}$ shrinks, as p increases.

Theorem 9: The measure of types choosing $x = 0$ is strictly increasing, and the measure of types choosing $x = \frac{1}{2}$ is decreasing (strictly if positive) in p. Moreover, the measure of types choosing $x = 0$ converges to zero as p approaches zero.

5.3 An additional observation

After circulating an earlier draft of this paper, we became aware of work by Dana et. al. [2006] and Brobert et. al. [2007], which shows that many dictators are willing to sacrifice part of the total prize to opt out of the game, provided that this decision is not revealed to recipients. Even though we did not develop our theory with these new experiments in mind, it provides an immediate explanation. Within our framework, opting out permits the dictator to retain a favorable image, $B(H)$ (the population average), while acting selfishly. In that sense, it is similar (but not identical) to choosing an action that could be attributable to nature. Not surprisingly, a positive mass of dictator types will take this option.

[12] Formally, it can be shown that a gap just above $x = 0$ definitely forms for p sufficiently close to unity, and definitely does not form for p sufficiently close to zero. However, since we do not attempt to test these implications, we omit a formal demonstration for the sake of brevity.

Formally, let's suppose that opting out permits the dictator to consume $y < x$. To construct an equilibrium, we divide the types into three segments, $[0, t_0]$, $(t_0, t_1]$, and $(t_1, \bar{t}]$, where $t_0 \leq t_1$. For $t \in [0, t_0]$, the dictator opts out; for $t \in (t_0, t_1]$, $Q(t) = S^{t_0}(t)$; and for $t \in (t_1, \bar{t}]$, $Q(t) = \frac{1}{2}$. This structure resembles that of a double pooling equilibrium, except that, instead of choosing x_0, the lowest segment opts out. Type t_0 must be indifferent between opting out and separating (where, as with double pooling equilibria, separation involves his first-best alternative, $x^*(t_0)$). Opting out provides a type t dictator with the utility level $F(y, B(H)) + tG\left(-\frac{y}{2}\right)$.Thus, the following indifference condition takes the place of (2):

$$F(y, B(H)) + t_0 G\left(-\frac{y}{2}\right) = U(x^*(t_0), t_0, t_0). \tag{7}$$

For $t_0 = \bar{t}$, the right-hand side of (7) is necessarily greater than the left; separating provides the dictator both with a better image and with a preferred distribution of consumption. If the penalty for opting out is sufficiently small (in other words, if y is sufficiently close to x), then, for $t_0 = 0$, the left-hand side of (7) is greater than the right-hand side; opting out provides a better image and virtually the same distribution of consumption. Therefore, with a small opt-out penalty, the solution to (7) is interior, which means that a positive mass of dictator types opts out.

6 Experimental Evidence

The theory described in the preceding sections not only accounts for previously observed behavioral patterns, but also generates additional testable implications. To our knowledge, no existing data shed light on the validity of those further implications. We therefore designed new experiments and collected fresh data. We focus on the theory's most simple and direct first-order implications: for the extended dictator game described in Section 5, increasing p should increase the mass of dictators who choose any given x_0 (close to zero) and reduce the mass who split the payoff equally. We test these implications by examining the effects of changes in p and x_0 on the distribution of dictators' choices.

6.1 Overview of the experiment

Subjects are divided into pairs, with partners and roles assigned randomly. Each pair splits a $20 prize. To facilitate interpretation, we will henceforth renormalize x, measuring it on a scale of 0 to 20. Thus, equal division corresponds to $x = 10$, rather than $x = 0.5$.

To heighten the effects of social image, dictators, recipients, and outcomes are publicly identified at the conclusion of the experiment. We manipulate the intensity of the resulting audience effect through the parameter p, which determines the audience's inclination to attribute an outcome to a dictator. We examine choices for four values of p ($p = 0$, 0.25, 0.5, and 0.75) and two values of x_0 ($x_0 = 0$, 1).

Identifying the distribution of voluntary choices for eight separate parameter combinations obviously requires a great deal of data. Suppose, for example, that we wish to have 30 observations of voluntary choices for each parameter combination. With each pair of subjects playing one game, we would require 1,000 subjects and $10,000 in prizes.[13] Therefore, our main challenge was to design experimental procedures that would allow us to gather sufficient data at reasonable cost.

One natural solution is to use the strategy method. In other words, we could ask each dictator to identify binding choices for a number of games, in each case conditional on nature permitting him to divide the prize, and then choose one game at random to determine the outcome. Unfortunately, this simple approach raises two important concerns.

First, in piloting the study, we discovered an apparent tendency for subjects to focus on *ex ante* fairness – that is, the equality of expected payoffs before nature's move – when choosing a strategy for a given game. If a dictator knows that nature will favor him when nature determines the split, the dictator will compensate by choosing a strategy that favors the recipient when the dictator determines the split. Thus, the strategies for a substantial fraction of dictators will prescribe gifts in excess of 50 percent of the prize. While this phenomenon raises some interesting questions concerning *ex ante* versus *ex post* fairness,

[13] We would require 30 pairs for each combination with $p = 0$, 40 pairs for each combination with $p = 0.25$, 60 pairs for each combination with $p = 0.5$, and 120 pairs for each combination with $p = 0.75$.

concerns for *ex ante* fairness are properly viewed as confounds in the context of our current investigation.

Second, the strategy method potentially introduces unintended and confounding audience effects. If a subject views the experimenter as part of the audience, his decisions may be influenced by the possibility that the experimenter will make inferences about the subject's character from his *strategy* rather than from the outcome. The experimenter potentially knows whether an outcome is attributable to nature or to the dictator, what the dictator would have chosen even if nature determined the outcome, and how the dictator would have played the game with other values of p and x_0. Our theory assumes that the relevant audience has none of this information.

We address these concerns through the following measures.

(1) We use the strategy method only to elicit choices for different games. We do not use it to elicit the subject's strategy for a particular game. For each game, the dictator is only asked to make a choice if he has been informed that his choice will govern the outcome. Thus, within each game, each decision is made *ex post* rather than *ex ante*, and there is no risk that the experimenter will draw inferences from portions of strategies that are never executed.

(2) We modify the extended dictator game described in Section 5 by making nature's choice symmetric. In our experiments, the dictator determines the split of the prize with probability $1 - p$. Nature sets the gift at x_0 with probability $p/2$ and at $20 - x_0$ with probability $p/2$. This symmetry largely eliminates the discrepancy between *ex ante* and *ex post* fairness, and should neutralize the tendency among dictators to compensate for any asymmetry in nature's choice. Notably, this modification does not alter any of the theoretical implications described in Section 5.[14]

(3) Our experimental procedures guarantee that no one associated with the experiment

[14] For the purpose of constructing an equilibrium, the mass at $20 - x_0$ can be ignored. It is straightforward to demonstrate that all types will prefer their equilibrium choices to this alternative, given it will be associated with the social image $B(H)$. They prefer their equilibrium choices to the action chosen by $\bar{t}$, and must prefer that choice to $20 - x_0$, because it provides more consumption, less inequality, and a better social image.

will be able to associate any dictator with his or her strategy. We make this fact evident to all subjects.

(4) Our experimental procedures also call subjects' attention to the fact that everyone present in the lab will associate each dictator with the outcome that emerges in the selected game. We thereby focus the subjects' attention on a particular audience and on the revelation of particular information (the outcome).

To avoid overwhelming subjects with large numbers of contingent choices, we created two distinct experimental conditions, designated "condition 0" and "condition 1." We set $x_0 = 0$ for condition 0 and $x_0 = 1$ for condition 1. Each pair of subjects is assigned to one of the two experimental conditions, and each dictator makes choices for all four values of p.

With this design, we identify the effects of x_0 from variation between subjects, and the effects of p from variation within subjects. When $p = 0$ we should observe the same distribution of choices for both conditions, including a spike at $x = 10,$ a 50-50 split. For $p = 0.25$, a second spike should appear, located at $x = 0$ for condition 0 and at $x = 1$ for condition 1. As we increase p to 0.50 and 0.75 the spikes at 10 should shrink and the spikes at x_0 should grow.

The subjects for our experiment were students enrolled in undergraduates economics courses at the University of Wisconsin–Madison in March and April 2006. We recruited a total of 120 subjects and divided them into two groups of 30 pairs, with one dictator and one recipient in each pair. One group of dictators made choices under condition 0, the other under condition 1. Due to unexpected attrition, we lost one pair for condition 1, leaving 29 pairs.

The next subsection describes our experimental protocol in greater detail. The terrain explored in this experiment provided some special challenges, and thus the design of our protocol required a bit of methodological innovation. Readers who are uninterested in experimental methods can skip directly to section 6.3, which presents results.

6.2 Details of the Protocol

The design of our experiment addresses four main challenges.[15] First, we must gather a substantial amount of data from a limited subject pool at reasonable cost, presumably through some variant of the strategy method. Second, we must induce subjects to focus on *ex post* fairness within each game. Third, we must establish a salient audience and minimize the likelihood that a subject will concern himself with the inferences of some spurious audience. Finally, we must make sure that subjects comprehend both the game's information structure and the odds that govern nature's choices. Dictators must understand that if they select $x = x_0$, the receiver will not be able to determine whether nature or the dictator chose the allocation. In this section, we describe how the experiment unfolded from the perspective of the subjects, and how particular design elements addressed these four challenges. Copies of the subjects' instructions are available from the authors.[16]

Each session included 20 subjects, all of whom were paid a $5 show-up fee. As they entered the experiment, participants were randomly assigned seats. Ten subjects sat on each side of the room. Those on one side were designated dictators, the others recipients. Each recipient was seated opposite the dictator with whom he or she was paired. Each pair was assigned a group number.

We began the experiment by asking each matched pair of subjects to stand and face each other, as in Bohnet and Frey (1999). They recited to each other the phrase, "Hello. I am in Group Number X . I am your partner." Subjects were told that one of them would be the "decision maker" (that is, the dictator), and that the other would be idle. Each dictator was given three envelopes. One, marked "blanks," contained nine decision sheets, described below. The other two, marked "completed" and "chosen," were empty.

[15] In the spirit of full disclosure, we acknowledge that we learned of the the second and fourth challenges through two "unsuccessful" pilot experiments. In the first it was clear that the design was excessively complex, and that subjects didn't understand either the probabilities governing nature's choices or the information structure. In the second, subjects overwhelmingly focused on *ex ante* fairness. The powerful and striking pull of *ex ante* fairness calls for further study, which we hope to pursue separately from this project.

[16] Go to http://econ.ucsd.edu/~jandreon/ or http://www.stanford.edu/~bernheim/.

Decision Sheet 7

My group number is 5

My private number is ______

Private Numbers 1 and 2 make a choice:

> "Divide $20: I allocate ______ to myself, and ______ to my partner."

Private Numbers 3 and 4: we are forcing you to make this choice:

> Write "forced" on this line: _______________
>
> If the coin flip is Heads:
> "Divide $20: I allocate __$20_ to myself, and __$0_ to my partner."
>
> If the coin flip is Tails:
> "Divide $20: I allocate __$0__ to myself, and __$20__ to my partner."

Features of the decision sheet we will report to your partner:

Odds of an intended decision:	2 in 4 (50%)
Odds of a forced decision:	2 in 4 (50%)

Figure 5: Example Decision Sheet

We then assigned to each dictator a "private number" using the following procedure. Dictators came to the front of the room one at a time, and each rolled a die until he obtained a number between 1 and 4. This private number was then written in ink at the top of each of the dictator's decision sheets (which already included the dictator's group number). The subject was instructed not to share this private number with anyone else.

Each decision sheet corresponded to a separate modified dictator game. We used separate sheets for separate games to underscore the notion that the dictator should consider each game in isolation. Figure 5 is an example of a decision sheet.

Notice that the method of allocating the $20 prize in Figure 5 depends on the dictator's private number. For some private numbers, the dictator determined the allocation of the prize by filling in the blanks in the following statement:[17]

"Divide $20: I allocate ______ to myself, and ______ to my partner."

For other private numbers, the dictator made no decision, instead submitting to a rule for determining the allocation. In that case, the dictator was asked to write "forced" on the decision sheet. Because each dictator wrote something on each sheet whether or not he or she chose the allocation, participants were unable to infer whether a particular decision was forced by watching the dictator.

For the decision sheet in Figure 5, the forced-choice rule was to allocate $20 to one partner and $0 to the other based on an unobserved coin flip. This rule corresponds to condition 0 ($x_0 = 0$). We replace these values with $19 and $1 for condition 1 ($x_0 = 1$). Note that nature's rule treats the dictator and recipient symmetrically. With this symmetric rule, we are more confident that no subject will, for instance, choose $x = 20$ to balance out the possibility that nature might have chosen $x_0 = 0$. Since nature is equally likely to be nice or nasty to the recipient, $x = 10$ remains the most natural fair allocation.

Notice also that the dictator makes choices *ex post* within each game – that is, after, nature determines whether the dictator controls the allocation for that game. This design feature has several advantages. First, it focuses the dictator's attention on *ex post* fairness, thereby reducing the likelihood that he will in some way attempt to compensate for nature's decision rule. Second, it eliminates one possible source of spurious audience effects (those arising from the experimenter's ability to observe choices that turn out to be irrelevant within a given game). Third, it underscores the fact that the dictator, unlike the audience, knows whether nature is responsible for the outcome.

We varied the value of p from one decision sheet to the next by changing the set of private numbers for which the dictator chose the allocation. This procedure made the odds

[17] Subjects were asked to check that the amounts summed to 20. All choices did.

of forced decisions transparent. For example, for the decision sheet in Figure 5, dictators with private numbers of 1 or 2 chose the allocation of the prize. Consequently, this decision sheet corresponds to a modified dictator game with parameter values $x_0 = 0$ and $p = 0.5$ (since that is the prior probability of a forced decision as perceived by any recipient). To assure transparency, we also listed the odds at the bottom of the decision sheet.

To guarantee that every dictator actually makes at least one allocation decision for every value of p, we used nine decision sheets. The nine sets of private numbers for which the dictator chose the allocation were $\{1, 2, 3, 4\}$, $\{1, 2, 3\}$, $\{2, 3, 4\}$, $\{1, 2\}$, $\{3, 4\}$, $\{1\}$, $\{2\}$, $\{3\}$, and $\{4\}$. With the set $\{1, 2, 3, 4\}$, all dictators chose allocations, so $p = 0$. With the sets $\{1, 2, 3\}$ and $\{2, 3, 4\}$, three out of four dictators chose allocations, so $p = 0.25$. With the sets $\{1, 2\}$ and $\{3, 4\}$, two out of four dictators chose allocations, so $p = 0.5$. Finally, with the sets $\{1\}$, $\{2\}$, $\{3\}$, and $\{4\}$, one out of four dictators chose an allocation, so $p = 0.75$.

Notice that we obtain one observation from each dictator for all values of p other than 0.25. For $p = 0.25$, we obtain one observation if the dictator's private number is 1 or 4 and two observations if that number is 2 or 3. In our experiment, 35 dictators actually made two decisions for $p = 0.25$. Of those, 29 made the same choice both times and 6 made different choices. When analyzing the data, we average the duplicative choices. Our results are not sensitive to this convention. Using the first, second, maximum, or minimum value leads to virtually identical conclusions.

Prior to each session, the order of the decision sheets was determined at random. However, all dictators within a single session filled out the sheets in the same order and at the same time. Once all private numbers had been assigned, dictators were instructed to remove the top decision sheet from the envelope marked "blanks." A copy of the sheet was displayed on an overhead projector so both dictators and recipients could see it. When subjects completed a form, they put it in the envelope marked "complete." Once all subjects completed a sheet, they were instructed to remove the next sheet from the "blanks" envelope.

After all nine forms were completed, the experimenter randomly selected the one that would be used to determine payments.[18] All dictators were instructed to remove the chosen decision sheet from the "complete" envelope and put it in the envelope marked "chosen." Both envelopes were sealed and the "chosen" envelopes were collected. Those envelopes were then handed to an assistant waiting outside the room. The assistant opened the envelopes in another room, determined payoffs, and placed earnings in "earnings envelopes" marked with the subjects' numbers. Without entering, the assistant returned the earnings envelopes to the original room, along with a summary of the outcomes. Since the assistant did not view any of the participants, it is doubtful that subjects regarded him as part of the audience.

The experimenter then wrote the final allocation for each pair on a board at the front of the room. The following example, which illustrates how outcomes would be displayed, was included in the subjects' instructions:

Chosen Decision Sheet: 8
Odds of an intended decision: 1 in 4 (25%)
Odds of a forced decision: 3 in 4 (75%)

Group 1	Decision maker - $10	Partner - $10
Group 2	Decision maker - $20	Partner - $0
Group 3	Decision maker - $9.10	Partner - $10.90
Group 4	Decision maker - $18	Partner - $2
Group 5	and so forth...	

The subjects' instructions also made it clear that, in this example, all participants would be able to infer that the dictators in groups 1, 3, and 4 surely determined the allocations for their groups, while the allocation for group 2 might have been chosen either by the dictator or by chance. To make sure that the subjects understood the game's information structure, the instructions included two other related examples.

At the conclusion of the session, subjects were handed their sealed earnings envelopes. After turning in the envelopes containing the unused decision sheets, they were free to leave. Subjects were assured at the outset of the session that the "complete" envelopes would be

[18] Randomization involved rolls of a 10-sided die. If a 10 appeared, the experimenter rolled the die again. Subjects observed this process.

opened much later, and that at no time would anyone who had been present in the room view any of their decision sheets.

While subjects were waiting for their payments, we asked them to fill out a questionnaire. This tested their understanding of the game by having them compute payoffs for both dictators and recipients in several examples, and state whether recipients could distinguish an intentional choice from a forced choice. All subjects—dictators and recipients—correctly answered the test questions, giving us confidence that the instructions were well understood.

As a check on our motivational assumptions, the questionnaire also asked about their goals and attitudes during the experiment. We discuss the responses to this questionnaire below.

6.3 Main Findings

In this section we evaluate the theory's central predictions. In subsequent sections we provide further analysis of the data at the subject level, and we discuss responses to the questionnaire. As we will see, the experimental data provides decisive support for the theory.

Figure 6 summarizes the distributions of dictators' voluntary choices in condition 0 ($x_0 = 0$) for each of the four values of p.[19] For ease of presentation, we group values of x into the following five categories: $x = 0$, $x = 1$, $2 \leq x \leq 9, x = 10,$ and $x > 10$. Figure 7 provides the same information for condition 1 ($x_0 = 1$).

The data depicted in Figures 6 and 7 provide striking confirmation of our theory's predictions. Look first at Figure 6 (condition 0). For $p = 0$ we expect a spike at $x = 10$. Indeed, 57 percent of dictators divided the prize equally. Consistent with results obtained from previous dictator experiments, a substantial fraction of subjects (30 percent) chose $x = 0$.[20]

[19] Although subjects were permitted to choose any division of the $20 prize, and although they were provided with hypothetical examples in which dictators chose allocations that involved fractional dollars, all chosen allocations involved whole dollars.

[20] For instance, the fraction of dictators who kept the entire prize was 35 percent in Forsythe et al. (1994) and 33% in Bohnet and Frey (1999). In contrast to our experiment, however, no dictators kept the entire prize in Bohnet and Frey's "two-way identification" condition. One potentially important difference is that

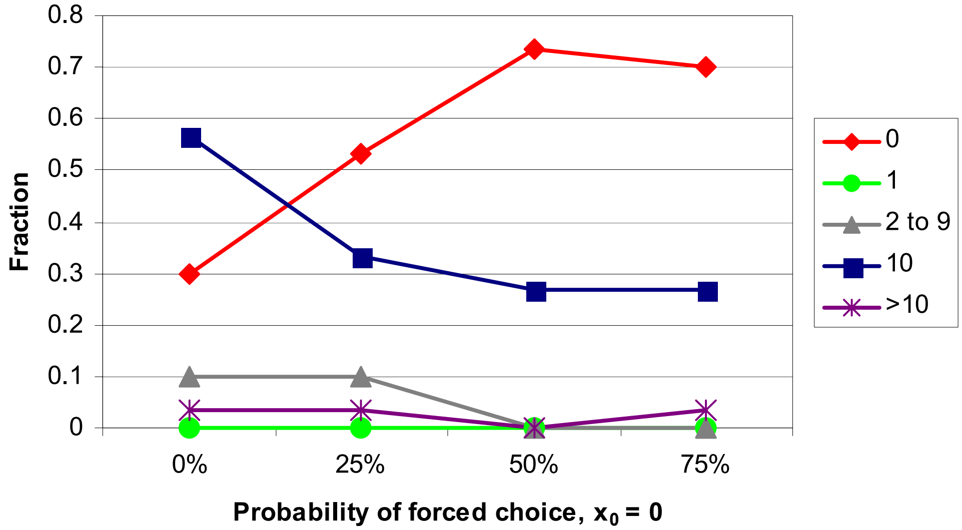

Figure 6: Distribution of Amounts Allocated to Partners, Condition 0

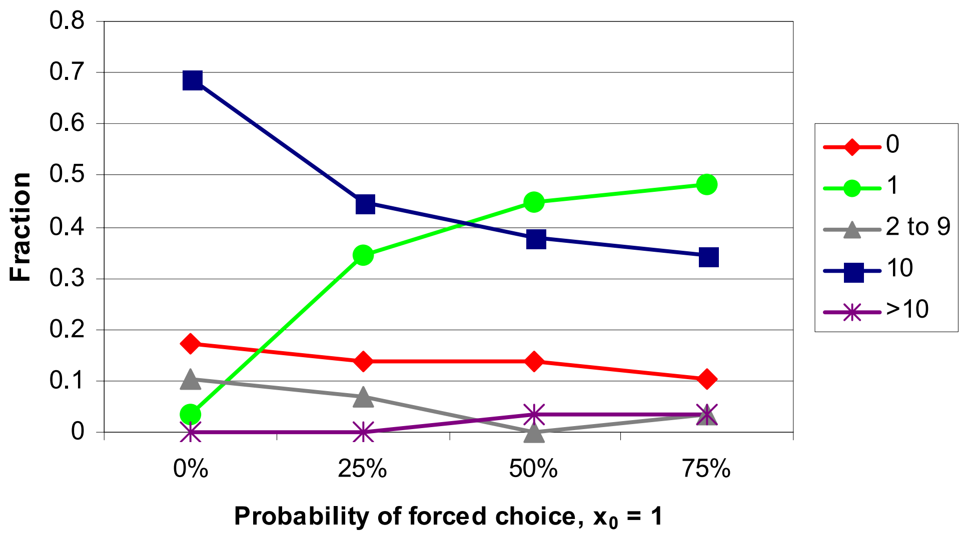

Figure 7: Distribution of Amounts Allocated to Partners, Condition 1

As we increase p we expect the spike at $x = 10$ to shrink and the spike at $x = 0$ to grow. That is precisely what happens. Note also that no subject chose $x = 1$ for any value of p.

Look next at Figure 7 (condition 1). Again, for $p = 0$ we expect a spike at $x = 10$. Indeed, 69 percent of dictators divide the prize equally, while 17 percent keep the entire prize ($x = 0$), and only 3% (one subject) chose $x = 1$. As we increase p the spike at $x = 10$ once again shrinks. In this case, however, a new spike emerges at $x = 1$. Specifically, as p increases to 0.75, the fraction of dictators choosing $x = 1$ rises steadily from 3 percent to 48 percent, while the fraction choosing $x = 10$ falls steadily from 69 percent to 34 percent. Notably, the fraction choosing $x = 0$ *falls* in this case from 17 percent to 10 percent. Once again, the effect of variations in p on the distribution of choices is dramatic, and exactly as predicted by the theory.

The statistical significance of these effects is demonstrated in Table 1, which reports estimates of two random-effects probit models. The specification in column (1) describes the probability of selecting $x = x_0$; the one in column (2) describes the probability of selecting $x = 10$, equal division. The explanatory variables include indicator variables for the values of the experimental parameters p (with $p = 0$ omitted) and x_0 (with $x_0 = 0$ omitted). In both cases, we report marginal effects at mean values, including the mean of the unobserved individual heterogeneity. For ease of presentation, we pool data from both conditions, although similar results hold on each condition separately.

In column (1), the coefficients on the indicator variables for all probabilities p are positive and highly statistically significant, indicating that an increase in p from 0 to any positive value raises the fraction of subjects choosing x_0. Moreover, the coefficient on the indicator for $p = 0.50$ is significantly higher than the one for $p = 0.25$ ($\chi^2 = 6.61$, $\alpha < 0.01$), as predicted. The coefficient on the indicator for $p = 0.75$ is virtually identical to the one for $p = 0.50$. Note that the coefficient on the indicator variable for $x_0 = 1$ is significant and negative. This may reflect the choices of a subset of subjects who are unconcerned with

Bohnet and Frey's subjects were all students in the same course, whereas our subjects were drawn from all undergraduates enrolled in economics courses at the University of Wisconsin, Madison.

social image, and are therefore more willing to join a pool at $x = 0$ than at $x_0 = 1$.

In column (2), the coefficients on the indicator variables for all probabilities p are negative and highly statistically significant, indicating that an increase in p from 0 to any positive value reduces the fraction of subjects choosing $x = 10$ (equal division). While the coefficients decline with p as predicted, the coefficient on the indicator for any given p is not statistically significantly different from its neighbor (for $p = 0.25$ vs. $p = 0.50$, $\chi^2 = 0.59$, and for $p = 0.50$ vs. $p = 0.75$, $\chi^2 = 0.04$). Notice that the coefficient on the indicator variable $x_0 = 1$ is not significantly different from zero in this regression.

TABLE 1
Random effects probit models:
marginal effects for regressions describing
(1) the probability of choosing $x = x_0$ and
(2) the probability of choosing $x = 10$ (equal division),
conditional on experimental parameters.†

	(1) $\Pr(x = x_0)$	(2) $\Pr(x = 10)$
$p = 0.25$	0.604**	−0.386**
	(0.112)	(0.111)
$p = 0.50$	0.801**	−0.466**
	(0.077)	(0.121)
$p = 0.75$	0.800**	−0.482**
	(0.077)	(0.122)
$x_0 = 1$	−0.524**	0.224
	(0.179)	(0.219)
Observations	236	236

†Standard errors in parentheses. Significance: ** at $\alpha < 0.01$.

As final check on the model's predictions, we compare choices across the two conditions for $p = 0$. Since the forced choice is irrelevant in those cases, we expect to find the same distribution of choices regardless of whether x_0 equals 0 or 1. As predicted, we find no significant difference between the two distributions (Mann-Whitney $z = 0.670, \alpha < 0.50$, Kolmogorov-Smirnov $k = 0.13, \alpha < 0.95$). The higher fraction of subjects choosing $x = 0$ in condition 0 (30 percent versus 17 percent) and the higher fraction choosing $x = 1$ in condition 1 (3 percent versus 0 percent) suggest a modest anchoring effect, but that pattern is also consistent with chance (comparing choices of $x = 0$, we find $t = 1.145, \alpha < 0.26$).

6.4 Subject-level analysis

One stark implication of our theory is that no subject should give away more than half of the prize. In previous experimental studies of the dictator game, violations of this prediction occasionally occur, but are relatively rare. The same is true of our experiment. For condition 0, there were three violations of this prediction, out of 139 total choices. One subject gave away \$15 when $p = 0$. A second subject gave away \$15 in one of two instances with $p = 0.25$ (but gave away \$10 in the other instance), and gave away \$11 when $p = 0.75$. For condition 1, there were only two violations of this prediction out of 134 total choices.[21] Both involved the same subject, who chose $x = 19$ with $p = 0.5$ and 0.75. When asked to explain her choices on the post-experiment questionnaire, this subject indicated that she alternated between giving \$1 and \$19 in order to "give me and my partner equal opportunities to make the same \$." Despite our precautions, this subject was clearly concerned with *ex ante* fairness.

Our theory also implies that, as p increases, a subject in condition 0 will not increase his gift, x. We find that 5 of 30 subjects violate this monotonicity prediction; for each of these subjects, there is a single violation. The same prediction holds for condition 1, with an important exception: according to theory, an increase in p could induce a subject to switch from $x = 0$ to $x = 1$. We find four violations of monotonicity for condition 1, but two involve switches from $x = 0$ to $x = 1$. One subject gave away nothing with $p = 0$, but gave \$1 for all higher values of p. A second gave away nothing for $p < 0.75$, but gave away \$1 for $p = 0.75$. Thus, problematic violations of monotonicity are relatively rare.

In sum, an examination of subject-level data strengthens our confidence in the theory. Allocations in excess of 50-50 are uncommon and subjects' choices are largely monotonic, as

[21] The total numbers of observations reported here exceeds the numbers reported in Tables 1 and 2 because here we do not average duplicative choices for $p = 0.25$.

the theory predicts.

6.5 Motivational Assumptions

The theory developed in this paper is based on two main assumptions concerning preferences: first, that people are fair-minded to varying degrees; second, that people like others to see them as fair. As a check on the validity of these assumptions, we included in the subjects' questionnaire several questions concerning attitudes and motives. We acknowledge that answers to such questions are potentially open to interpretation and rarely suffice to prove or disprove an economic theory. However, since the motives envisioned in our model are nonstandard, we feel it is useful to supplement our examination of indirect behavioral evidence (discussed above) with direct evidence concerning objectives. Here we focus on three questions that pertain to the model's assumptions.

Subjects were presented with a list of possible objectives and asked to indicate the importance of each on a scale of 1 to 5, with 1 signifying "not important" and 5 signifying "very important." The list included the following objectives (where the use of 19 or 20 depended on the condition):

a) Making the most money I could.

b) Being generous to my partner.

c) Not getting caught when I chose 20 (19) for me.

The importance of objective (a) should correlate with selfishness, while the importance of objective (b) should correlated with altruism or fairness. We would expect those who endorse (a) to be more likely to choose $x = x_0$, and those who endorse (b) to be more likely to choose $x = 10$. Statement (c) acknowledges a desire to mask intentions by disguising selfish actions. Those who endorse (c) should be more likely to select $x = x_0$ and less likely to choose $x = 10$, but only when $p > 0$.

To check these hypotheses we estimated additional random-effects probit models, which we report in Table 2. Column (1) shows that endorsing (a), the desire to make money, is strongly positively related to choosing $x = x_0$. Endorsing (c), the desire to hide selfishness, is strongly positively related to choosing $x = x_0$ when $p > 0$, but not when $p = 0$, exactly as our theory predicts. Column (2) shows that endorsing (b), the desire to be generous, is significantly related to choosing $x = 10$. Endorsing (c) is significantly negatively related to choosing $x = 10$ when $p > 0$, but not when $p = 0$, again exactly as our theory predicts. Thus, the patterns in these regressions are consistent with our underlying behavioral assumptions.

TABLE 2

Random-effects probit models: marginal effects for regressions describing (1) the probability of choosing $x = x_0$,and (2) the probability of choosing equal division ($x = 10$), conditional on self-reported motivations, and interactions with an indicator for $p > 0$.[†]

	(1) $\Pr(x = x_0)$	(2) $\Pr(x = 10)$
a. Making money	0.351**	-0.161
	(0.125)	(0.109)
b. Being generous	-0.005	0.232*
	(0.100)	(0.092)
c. Not getting caught	0.011	0.064
	(0.082)	(0.065)
$a.\times 1(p > 0)$	0.101	-0.138*
	(0.070)	(0.067)
$b.\times 1(p > 0)$	-0.173	0.152*
	(0.099)	(0.071)
$c.\times 1(p > 0)$	0.296**	-0.180*
	(0.095)	(0.081)
Observations	236	236

[†]Standard errors in parentheses. Significance: ** at $\alpha < 0.01$, * at $\alpha < 0.05$

7 Concluding Comments

We have proposed and tested a theory of behavior in the dictator game that is predicated on two critical assumptions: first, that people are fair-minded to varying degrees; second, that people like others to see them as fair. We have demonstrated that this theory accounts for previously unexplained aspects of observed behavior. It also has ancillary implications that are both sharp and testable. Because no existing data shed light on the validity of those further implications, we designed a new experiment and collected fresh data. The data obtained from our experiment strongly support the theory.

Narrowly interpreted, this study enriches our understanding of behavior in the dictator game. More generally, it provides a theoretical framework that potentially accounts for the prevalence of the equal division norm in a wide range of settings. Perhaps most significantly, it underscores both the importance and feasibility of studying audience effects with theoretical precision, and it suggests that economists can gain an understanding of those effects by modeling them as signaling phenomena. Formal models of audience effects, like the one proposed here, can potentially help economists design better experiments and interpret experimental results more accurately. To the extent audience effects are pervasive in real economic choices (as we claim), such models will also prove useful more generally.

References

[1] Agrawal, Pradeep. 2002. "Incentives, Risk, and Agency Costs in the Choice of Contractual Arrangements in Agriculture." *Review of Development Economics* 6, 460-77.

[2] Andreoni, J. and R. Petrie. 2004. "Public Goods Experiments Without Confidentiality: a Glimpse into Fund-raising." *Journal of Public Economics* 88, 1605-1623.

[3] Bagwell, L. S. and B. D. Bernheim. 1996. "Veblen Effects in a Theory of Conspicuous Consumption." *American Economic Review* 86(3), 349-373.

[4] Bernheim, B. D. 1994. "A Theory of Conformity." *Journal of Political Economy* 102(5), 841-877.

[5] Bernheim, B. D. and S. Severinov. 2003. "Bequests as Signals: An Explanation for the Equal Division Puzzle." *Journal of Political Economy* 111, 733-764.

[6] Bloom, D.E. 1986. "Empirical Models of Arbitrator Behavior under Conventional Arbitration." *Review of Economics and Statistics* 68, 578-85.

[7] Bohnet, I. and B. S. Frey. 1999. "The Sound of Silence in Prisoner's Dilemma and Dictator Games." *Journal of Economic Behavior & Organization* 38, 43-57.

[8] Broberg, T., T. Ellingsen, and M. Johannesson. 2007. "Is Generosity Involuntary?" *Economic Letters* 94, 32-37.

[9] Camerer, C. F. 1997. "Progress in Behavioral Game Theory." *Journal of Economic Perspectives* 11, 167-188.

[10] Camerer, C. F. 2003. *Behavioral Game Theory: Experiments in Strategic Interaction.* Princeton University Press, Princeton.

[11] Cho, I.-K. and D. Kreps. 1987. "Signaling Games and Stable Equilibria." *Quarterly Journal of Economics* 102, 179-221.

[12] Cho, I.-K. and J. Sobel. 1990. "Strategic Stability and Uniqueness in Signaling Games." *Journal of Economic Theory* 50, 381-413.

[13] Dana, J., D. M. Cain, and R. 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, 193-201.

[14] Dasgupta, S. and Z. Tao. 1998. "Contractual Incompleteness and the Optimality of Equity Joint Ventures." *Journal of Economic Behavior and Organization* 37, 391-413.

[15] De Weaver, M. and J. Roumasset. 2002. "Risk Aversion as Effort Incentive: A Correction and Prima Facie Test of the Moral Hazard Theory of Share Tenancy." *Economics Bulletin* 15, 1-16.

[16] Ellingsen, T. and M. Johannesson. "Pride and Prejudice: The Human Side of Incentive Theory." Mimeo, Stockholm School of Economics, 2006.

[17] Fehr, E. and K. Schmidt. 1999. "A Theory of Fairness, Competition, and Cooperation." *Quarterly Journal of Economics* 114, 817-68.

[18] Forsythe, R., J. Horowitz, N. Savin, and M. Sefton. 1994. "The Statistical Analysis of Experiments with Simple Bargaining Games." *Games and Economic Behavior* 6, 347-69.

[19] Geanakoplos, J., D. Pearce, and E. Stacchetti. 1989. "Psychological Games and Sequential Rationality." *Games and Economic Behavior* 1, 60-79.

[20] Harbaugh, W. 1998. "The Prestige Motive for Making Charitable Transfers." *American Economic Review* 88, 277– 282.

[21] Hauswald, R. and U. Hege. 2003. "Ownership and Control in Joint Ventures: Theory and Evidence." C.E.P.R. Discussion Paper 4056.

[22] Hoffman, E., K. McCabe, and V. Smith. 1996. "Social Distance and Other-Regarding Behavior in Dictator Games." *American Economic Review* 86, 653-60.

[23] Menchik, P.L. 1980. "Primogeniture, Equal Sharing„ and the U.S. Distribution of Wealth." *Quarterly Journal of Economics* 94, 299-316.

[24] Manning, A. 2007. "Respect." Mimeo, London School of Economics.

[25] Menchik, P.L. 1988. "Unequal Estate Division: Is It Altruism, Reverse Bequests, or Simply Noise?" In *Modelling the Accumulation and Distribution of Wealth.* D. Kessler and A. Masson (eds.). New York: Oxford Univ. Press.

[26] Rege, M. and K. Telle. 2004. "The Impact of Social Approval and Framing on Cooperation in Public Good Situations" *Journal of Public Economics* 88, 1625-1644.

[27] Soetevent, A.R. 2005. "Anonymity in Giving in a Natural Context—A Field Experiment in 30 Churches." *Journal of Public Economics,* 89, 2301-2323.

[28] Tadelis, S. 2007. "The Power of Shame and the Rationality of Trust." Mimeo, Hass School of Business, U. C. Berkeley.

[29] Veugelers, R. and K. Kesteloot. 1996. "Bargained Shares in Joint Ventures among Asymmetric Partners: Is the Matthew Effect Catalyzing?" *Journal of Economics* 64, 23-51.

[30] Wilhelm, M.O. 1996. "Bequest Behavior and the Effect of Heirs' Earnings: Testing the Altruistic Model of Bequests." *American Economic Review* 86, 874-92.

Appendix

Proof of Theorem 1: Consider the problem

$$\max_{x\in[0,1]} F(1-x,t)+tG\left(x-\frac{1}{2}\right)$$

The solution involves one of the following three conditions:

$$tG'\left(x-\frac{1}{2}\right)=F_1(1-x,t) \tag{8}$$

$$tG'\left(-\frac{1}{2}\right)<F_1(1,t) \quad \text{(in which case } x=0\text{)} \tag{9}$$

$$tG'\left(\frac{1}{2}\right)>F_1(0,t) \quad \text{(in which case } x=1\text{)} \tag{10}$$

(1) We know that F_1 is strictly positive, and that $G'(z)$ is non-positive when $z\geq 0$. Clearly, (10) is never satisfied, and (8) cannot be satisfied for $x\geq\frac{1}{2}$.

(2) Implicitly differentiating (8), we obtain

$$\frac{dx}{dt}=-\frac{G'\left(x-\frac{1}{2}\right)-F_{12}(1-x,t)}{tG''\left(x-\frac{1}{2}\right)+F_{11}(1-x,t)} \tag{11}$$

With $x<\frac{1}{2}$ and separability of F, the numerator is strictly positive. By strict concavity, the denominator is strictly negative. Thus, the entire term is strictly positive. Note that, if $x=0$ is optimal for some t, then, by (9), it is also optimal for smaller t provided that $F_{12}\leq 0$.

(3) Let

$$t^*=\min\left\{\frac{f}{G'\left(-\frac{1}{2}\right)},\bar{t}\right\}$$

(where f was defined before as the lower bound on F_1). For all $t<t^*$, (9) is satisfied by construction.

(4) Suppose this is false. Then there exists a sequence $<t_k>_{k=0}^{\infty}$ increasing without bound and some $x_0<\frac{1}{2}$ such that, for all k, $x^*(t_k)<x_0$. In that case, $t_kG'\left(x^*(t_k)-\frac{1}{2}\right)>t_kG'\left(x_0-\frac{1}{2}\right)>0$, so as k goes to infinity $t_kG'\left(x^*(t_k)-\frac{1}{2}\right)$ increases without bound. Since

F_1 is bounded, this means that neither (8) or (9) can be satisfied. Since a maximum exists, and since (10) cannot be satisfied either, this is a contradiction. □

We now prove two general lemmas concerning (pure strategy) signaling equilibria, which we use in subsequent proofs.

Lemma 1 In equilibrium, $G\left(X(t) - \frac{1}{2}\right)$ is weakly increasing in t. (In other words, more fair types choose allocations that are weakly more fair.)

Proof: Consider two types, t and t' with $t < t'$. Suppose type t chooses x earning image m, while t' chooses x' earning image m'. Let $f = F(1-x, m)$, $f' = F(1-x', m')$, $g = G\left(x - \frac{1}{2}\right)$, and $g' = G\left(x' - \frac{1}{2}\right)$. Mutual non-imitation requires

$$f' + t'g' \geq f + t'g$$

and

$$f' + tg' \leq f + tg$$

Subtracting the second expression from the first yields

$$(g' - g)(t' - t) \geq 0$$

Since $t' - t > 0$, it follows that $g' - g \geq 0$. □

Corollary 1 Let $T^L = \{t \mid Q(t) \leq \frac{1}{2}\}$, and $T^H = \{t \mid Q(t) \geq \frac{1}{2}\}$. The equilibrium action function $Q(t)$ is weakly increasing in t on T^L and weakly decreasing in t on T^H.

Lemma 2 Suppose that $Q(t) > \frac{1}{2}$. Define x' as the solution to $G\left(x' - \frac{1}{2}\right) = G\left(Q(t) - \frac{1}{2}\right)$ (if G is not symmetric around $\frac{1}{2}$, x' may not exist, but the statement of the lemma still applies) Then, for all $t' > t$, $Q(t') \in \{x', Q(t)\}$. In other words, if type t chooses to give away more than half of the prize, then all higher types must either choose the same action as t, or choose to give away an equally fair amount less than $\frac{1}{2}$.

Proof: According to lemma 1, $G\left(Q(t')-\frac{1}{2}\right) \geq G\left(Q(t)-\frac{1}{2}\right)$. To prove the lemma, we show that this must hold with equality. Suppose on the contrary that the inequality is strict for some t'. Since G is single-peaked, $1-Q(t') > 1-Q(t)$. We claim that, in addition, $B\left(\Phi_{Q(t')}\right) \geq B\left(\Phi_{Q(t)}\right)$. Let $t^0 = \inf\{\tau \mid Q(\tau) = Q(t')\}$. Plainly, $B\left(\Phi_{Q(t')}\right) \geq t^0$. From lemma 1, we know that, for all $t'' > t^0$, $Q(t'') \neq Q(t)$. Thus, $B\left(\Phi_{Q(t)}\right) \leq t^0$, which establishes the claim. It follows that $F\left(1-Q(t'), B\left(\Phi_{Q(t')}\right)\right) > F\left(1-Q(t), B\left(\Phi_{Q(t)}\right)\right)$. But then all types, including t, prefer $Q(t')$ to $Q(t)$, a contradiction. □

Corollary 2 There is at most one value of x greater than $\frac{1}{2}$ chosen in any equilibrium.

Proof of Theorem 2: (1) First we show that $S_{r,z}(t) > x^*(t)$ for $t > r$ and $t-r$ sufficiently small. If $z > x^*(r)$, this is obvious. Suppose $z = x^*(r)$. If $x^*(r)$ is characterized by (9), then $x^*(r) = 0$, and it's easy to check that $S'_{r,z}(r) > 0$ (the numerator is positive, and the denominator is negative because the best choice for r, ignoring social image, is strictly less than zero). Since (9) also implies that $x^*(r) = 0$ for t close to r, we have the desired conclusion. The other possibility is that $x^*(r)$ is characterized by (8). In that case, $S'_{r,z}(r)$ is positive infinity. Inspection of (11) reveals that $\left.\frac{dx^*(t)}{dt}\right|_{t=r}$ is finite, so again the desired conclusion follows directly.

Next we argue that $S_{r,z}(t) > x^*(t)$ for all $t > r$. Suppose not. Then, since the solution $S_{r,z}$ must be continuous, there is some t' such that $S_{r,z}(t') = x^*(t')$ and $S_{r,z}(t) > x^*(t)$ for $t < t'$. Consider any monotonic sequence $t_k \uparrow t'$. As k increases, the denominator in the expression for $S'_{r,z}(t_k)$ approaches zero from below while the numerator converges to a strictly positive number, implying that $S'_{r,z}(t_k)$ increases without bound. In contrast, given our assumptions about F and G, the derivative of $x^*(t)$ is bounded within any neighborhood of t' (see equation (11) – G' achieves a maximum on $\left[-\frac{1}{2}, +\frac{1}{2}\right]$, G'' achieves a strictly negative maximum on $\left[-\frac{1}{2}, +\frac{1}{2}\right]$, F_{12} achieves a strictly negative minimum on $[0,1]\times[t'-\varepsilon, t'+\varepsilon]$, and F_{11} achieves a strictly negative maximum on $[0,1]\times[t'-\varepsilon, t'+\varepsilon]$). But then $S_{r,z}(t) - x^*(t)$

must increase over some interval (t'', t') (with $t'' < t'$). This implies $S_{r,z}(t') - x^*(t') > S_{r,z}(t'') - x^*(t'') > 0$, which contradicts $S_{r,z}(t') - x^*(t') = 0$.

(2) The numerator of (1) is always strictly positive and, in light of part (1), the denominator is always strictly negative for $t > r$. The negative sign in front of the fraction makes the derivative positive. As discussed in the proof of part (1), $S'_{r,z}(r)$ may be infinite.

(3) Consider t' and t'' with $S_{r,z}(t')$, $S_{r,z}(t'') \leq \frac{1}{2}$. Assume that $t' < t''$. Then

$$U\left(S_{r,z}(t''), t'', t'\right) = U\left(S_{r,z}(t'), t', t'\right) + \int_{t'}^{t''} \frac{dU\left(S_{r,z}(t), t, t'\right)}{dt} dt \tag{12}$$

But

$$\begin{aligned}
\int_{t'}^{t''} \frac{dU\left(S_{r,z}(t), t, t'\right)}{dt} dt &= \int_{t'}^{t''} \left\{ \left[t'G'\left(S_{r,z}(t) - \frac{1}{2}\right) - F_1\left(1 - S_{r,z}(t), t\right) \right] S'_{r,z}(t) + F_2\left(1 - S_{r,z}(t), t\right) \right\} dt \\
&< \int_{t'}^{t''} \left\{ \left[tG'\left(S_{r,z}(t) - \frac{1}{2}\right) - F_1\left(1 - S_{r,z}(t), t\right) \right] S'_{r,z}(t) + F_2\left(1 - S_{r,z}(t), t\right) \right\} dt \\
&= 0
\end{aligned}$$

where the inequality follows from $S_{r,z}(t) < \frac{1}{2}$, and where the final equality follows from (1). From (12), this implies

$$U\left(S_{r,z}(t''), t'', t'\right) < U\left(S_{r,z}(t'), t', t'\right)$$

The argument for $t'' < t'$ is symmetric.

(4) Assume that the claim is false. Since the solution to the differential equation is continuous, we must have $S_{r,z}(t) < \frac{1}{2}$ for all $t \geq r$. Note that, since F is unbounded in m, $U\left(\frac{1}{2}, m, r\right)$ exceeds $U\left(z, r, r\right) = U(S_{r,z}(r), r, r)$ for m sufficiently large. Since $S_{r,z}(t)$ lies between t's first best choice and $\frac{1}{2}$, since t's first-best choice approaches $\frac{1}{2}$ as t gets large (Theorem 1 part (4)), and since F_1 is bounded, we know that $U\left(S_{r,z}(t), t, r\right)$ converges to $U(\frac{1}{2}, t, r)$, which in turn exceeds $U(S_{r,z}(r), r, r)$ for t sufficiently large. But this contradicts part (3). (If F is assumed to be separable, the proof is simpler.)

(5) If $z > z'$, then $S_{r,z}(r) > S_{r,z'}(r)$. We wish to show that $S_{r,z}(t) > S_{r,z'}(t)$ for all $t > 0$. If this is not the case, then there exists some $t' > r$ for which $S_{r,z}(t) = S_{r,z'}(t)$. But, for

standard reasons, two trajectories with different initial conditions cannot intersect. Continuity in r and z follows from standard properties of the solutions of differential equations. $\square$

Proof of Theorem 3: First we argue that, in any separating equilibrium with action function Q, we must have $Q(t) < \frac{1}{2}$ for all $t < \bar{t}$. We prove this in two steps. The first step is to show $Q(t) \leq \frac{1}{2}$ for all $t < \bar{t}$. Corollary 2 tells us that there is at most one value of x greater than $\frac{1}{2}$ chosen in any signaling equilibrium. For a separating equilibrium, this means that at most one type, call it t', that chooses an action greater than $\frac{1}{2}$. Assume $t' < \bar{t}$. Consider $t'' \in \left(t', \bar{t}\right)$. We know that $Q(t'') \leq \frac{1}{2} < Q(t')$, so $F(1 - Q(t''), t'') > F(1 - Q(t'), t')$. From Lemma 1, $G\left(Q(t'') - \frac{1}{2}\right) \geq G\left(Q(t') - \frac{1}{2}\right)$. But then t' would imitate t'', a contradiction. The second step is to rule out $Q(t) = \frac{1}{2}$ for all $t < \bar{t}$. Assume on the contrary that $Q(t') = \frac{1}{2}$ for some $t' < \bar{t}$. By Lemma 1, $Q(t'') = \frac{1}{2}$ for all $t'' > t$. But then Q is not a separating function. (From Corollary 2, it then follows that any separation function is monotonic. Through a more elaborate argument, we can also show that $Q(\bar{t}) \leq \frac{1}{2}$. Neither of these conclusions are necessary for what follows.)

The argument in the preceding paragraph tells us that, in a separating equilibrium, all types $t \in [0, \bar{t})$ choose actions in the "standard" half of the action set where the single-crossing property is satisfied. Standard arguments then imply that a separating action function must solve (1) for some initial condition (Mailath [1987]). In other words, the separating action function must be $S_{0,z}(t)$ for some z. Moreover, from the preceding paragraph, we know we must have $S_{0,z}(\bar{t}) \leq \frac{1}{2}$. Since $S_{0,z}(\bar{t})$ is increasing in z, it is possible to construct a separating equilibrium iff $S_0(\bar{t}) \leq \frac{1}{2}$. Since $S_0(t)$ is strictly monotonic, this is equivalent to the statement that $\bar{t} \leq t^*$. $\square$

Proof of Theorem 4: Define $\psi(t_0)$ as the solution to

$$U(S_0(t_0), t_0, t_0) = U\left(\frac{1}{2}, \psi(t_0), t_0\right)$$

Note that we can rewrite the equilibrium condition as either (1) $\psi(t_0) = B\left(H_{t_0}\right)$, or (2)

$\psi(0) < B(H)$.

Let's start by looking at the properties of ψ. Under our assumptions, $\psi(t_0)$ always exists, is unique, and exceeds t_0 (since $x^*(t_0) < S_0(t_0) < \frac{1}{2}$, we know that $U(S_0(t_0), t_0, t_0) > U\left(\frac{1}{2}, t_0, t_0\right)$). Also, $\psi(t^*_{0,0}) = t^*_{0,0}$. We claim that $\psi(t_0)$ is also strictly decreasing and continuous in t_0. Continuity follows because all the pertinent functions (including S) are continuous. To see that the function is strictly decreasing, consider any t' and t'' with $t' > t''$. We know that

$$F\left(1 - S_0(t'), t'\right) - F\left(\frac{1}{2}, \psi(t')\right) = t'\left[G\left(\frac{1}{2}\right) - G\left(S_0(t') - \frac{1}{2}\right)\right]$$

and

$$F\left(1 - S_0(t''), t''\right) - F\left(\frac{1}{2}, \psi(t'')\right) = t''\left[G\left(\frac{1}{2}\right) - G\left(S_0(t'') - \frac{1}{2}\right)\right]$$

$>$From the first expression, we know that

$$F\left(1 - S_0(t'), t'\right) - F\left(\frac{1}{2}, \psi(t')\right) > t''\left[G\left(\frac{1}{2}\right) - G\left(S_0(t') - \frac{1}{2}\right)\right]$$

Subtracting the third expression from the second yields

$$\begin{aligned}&\left[F\left(1 - S_0(t''), t''\right) - F\left(1 - S_0(t'), t'\right)\right] + \left[F\left(\frac{1}{2}, \psi(t')\right) - F\left(\frac{1}{2}, \psi(t'')\right)\right] \\ < \;& t''\left[G\left(S_0(t') - \frac{1}{2}\right) - G\left(S_0(t'') - \frac{1}{2}\right)\right]\end{aligned}$$

But we also know (by the non-imitation constraint) that

$$\left[F\left(1 - S_0(t''), t''\right) - F\left(1 - S_0(t'), t'\right)\right] \geq t''\left[G\left(S_0(t') - \frac{1}{2}\right) - G\left(S_0(t'') - \frac{1}{2}\right)\right]$$

These two inequalities can hold only if

$$F\left(\frac{1}{2}, \psi(t')\right) - F\left(\frac{1}{2}, \psi(t'')\right) < 0$$

which requires $\psi(t'') > \psi(t')$.

Now let's think about the properties of $B(H_{t_0})$. Given our assumptions on B, it's plainly increasing and continuous in t_0. Moreover, $B\left(H_{t^*_{0,0}}\right) > t^*_{0,0}$.

From the preceding, we know that $B(H_{t_0}) - \psi(t_0)$ is strictly increasing and continuous in t_0.

Now let's look for equilibria. Assume $\bar{t} > t^*_{0,0}$. If $B(H) - \psi(0) \geq 0$, then we automatically have an equilibrium (all types choose $x = \frac{1}{2}$), and it's plainly unique because $B(H_{t_0}) - \psi(t_0) > 0$ for all $t_0 > 0$. Suppose instead that $B(H) - \psi(0) < 0$. We know that $B\left(H_{t^*_{0,0}}\right) - \psi(t^*_{0,0}) > 0$. Thus, there exists a unique value of t_0, necessarily on the open interval $(0, t^*_{0,0})$ (which gives us $S_0(t_0) < \frac{1}{2}$), for which $B(H_{t_0}) - \psi(t_0) = 0$.

To complete the description of an equilibrium, we need to supply beliefs for actions not chosen in equilibrium (all those in $[S(t_0), \frac{1}{2})$ and $(\frac{1}{2}, 1]$). We assume these actions are attributed to t_0 (when every type joins the pool, type zero). It is easy to check that no type would select any of these actions.

Now suppose $\bar{t} < t^*_{0,0}$. Then , for t_0 sufficiently close to $\bar{t}$, $B(H_{t_0})$ is very close to t_0. But since $S_0(t_0)$ is bounded away from $\frac{1}{2}$, $\psi(t_0)$ is bounded away from t_0. Therefore, $B(H_{t_0}) - \psi(t_0) < 0$. Since this function is strictly increase in t_0, we know that $B(H_t) - \psi(t) < 0$ for all $t \in [0, \bar{t}]$. But this means that there is no central pooling equilibrium.

For the case of $\bar{t} = t^*_{0,0}$, we have $B(H_{t^*_{0,0}}) - \psi(t^*_{0,0}) = 0$, so the central pool is empty (it consists of $(\bar{t}, \bar{t}]$). One can also think of this as a case with a degenerate pool at $x = \frac{1}{2}$ (it just consists of $\bar{t}$). $\square$

Proof of Theorem 5: We establish this result through a series of lemmas.

Lemma 3 Consider an equilibrium in which some nondegenerate set of types (a pool) selects some action other than $\frac{1}{2}$. This equilibrium does not satisfy the D1 criterion.

Proof: Suppose there is a pool that selects an action $x' \neq \frac{1}{2}$. Select some type t' belonging to the pool such that $t' > B(\Phi_{x'})$. We claim that, for any action x'' with $G\left(x'' - \frac{1}{2}\right) > G\left(x' - \frac{1}{2}\right)$, $B(\Phi_{x''}) \geq t'$. The lemma follows immediately from the claim because t' can then choose an action slightly closer to $\frac{1}{2}$ than x' and obtain a discontinuous increase in payoff from a discrete jump in social image (any other effect on utility can be made arbitrarily small by taking the new choice sufficiently close to x').

There are two cases to consider.

Case #1: x'' is chosen by some type in equilibrium. In that case, the claim follows immediately from Lemma 1, which tells us that no type lower than t' chooses x''.

Case #2: x'' is not chosen by some type in equilibrium. Consider any $t'' < t'$. We argue that the set of inference for which t'' would select x'' rather than its equilibrium choice is strictly smaller than the set of inferences for which t' would select x'' rather than its equilibrium choice. By the D1 criterion, this means that the inference upon seeing x'' places no weight on any type $t'' < t'$, from which the claim follows directly.

Consider any m such that

$$F(1-x'',m) - F\left(1-Q(t''), B\left(\Phi_{Q(t'')}\right)\right) \geq t''\left[G\left(Q(t'')-\frac{1}{2}\right) - G\left(x''-\frac{1}{2}\right)\right]$$

(which means t'' is weakly willing to pick x''). We also know (from non-imitation) that

$$F\left(1-Q(t''), B\left(\Phi_{Q(t'')}\right)\right) - F(1-x', B(\Phi_{x'})) \geq t''\left[G\left(x'-\frac{1}{2}\right) - G\left(Q(t'')-\frac{1}{2}\right)\right]$$

Adding these inequalities yields

$$F(1-x'',m) - F(1-x', B(\Phi_{x'})) \geq t''\left[G\left(x'-\frac{1}{2}\right) - G\left(x''-\frac{1}{2}\right)\right]$$

Since, by assumption, $t'' < t'$ and $G\left(x''-\frac{1}{2}\right) > G\left(x'-\frac{1}{2}\right)$, we have

$$F(1-x'',m) - F(1-x', B(\Phi_{x'})) > t'\left[G\left(x'-\frac{1}{2}\right) - G\left(x''-\frac{1}{2}\right)\right]$$

Since $Q(t') = x'$, this means that t' strictly prefers to choose x'', given the inference m. So t' is willing to choose x'' for a strictly larger set of inferences, as claimed. $\square$

Lemma 4 In any equilibrium satisfying the D1 criterion, type $t = 0$ selects either $x = 0$ or $x = \frac{1}{2}$.

Proof: Suppose $Q(0) \notin \{0, \frac{1}{2}\}$. By lemma 3, $\Phi_{Q(0)}$ places probability one on type 0. But then

$$U(0, B(\Phi_0), 0) \geq U(0,0,0) > U\left(Q(0), B\left(\Phi_{Q(0)}\right), 0\right)$$

which contradicts the assumption that $Q(0)$ is the optimal choice for type $t = 0$. $\square$

Now we prove that the D1 criterion excludes all equilibria other than the ones mentioned in the statement of the theorem. Assume first that an equilibrium satisfying the D1 criterion includes a pool at $\frac{1}{2}$. Lemmas 1 and 2 together imply that $Q(t) \leq \frac{1}{2}$ for all $t \in [0, \bar{t}]$. By Corollary 1, the action function is weakly monotonic, so by Lemma 3 either all types choose $x = \frac{1}{2}$, or we can partition the types into two sets, $(t_0, \bar{t}]$, who choose $\frac{1}{2}$, and $[0, t_0]$, each of whom chooses a different action. (Note again that t_0 can be placed in either group, and that the choice is inconsequential because all types have measure zero; by convention, we place t_0 in the separating group.)

Suppose all types choose $x = \frac{1}{2}$. Then

$$U\left(\frac{1}{2}, B(H), 0\right) > U\left(0, B\left(\Phi_0\right), 0\right) \geq U(0,0,0)$$

(where the strict inequality follows from the fact that we resolve indifference in favor of lower choices), since otherwise type zero would choose $x = 0$. But this is then a central pooling equilibrium.

Now suppose that some types do not choose $x = \frac{1}{2}$. For types in $[0, t_0]$, we have a completely standard problem of separation (because we know their actions must lie in the standard half of the choice set). Standard arguments then imply that their actions must solve (1), and Lemma 4 implies that the initial condition is $S(0) = 0$. Thus, for types in $[0, t_0]$, actions are given by $S_0(t)$. Moreover, type t_0 must be indifferent between $\left(\frac{1}{2}, B\left(H_{t_0}\right)\right)$ and $(S_0(t_0), t_0)$ – if t_0 strictly preferred the first alternative then, by continuity, types slightly lower than t_0 would strictly prefer to enter the central pool; if t_0 strictly preferred the second alternative then, by continuity, types slightly higher than t_0 would strictly prefer to exit the pool by imitating t_0. Accordingly, the configuration is again a central pooling equilibrium.

Now assume that an equilibrium satisfying the D1 criterion does not include a pool at $\frac{1}{2}$. By Lemma 3, we know it's a separating equilibrium. In the proof of Theorem 3, we showed that a separating equilibrium action function must equal $S_z(t)$ for $t \in [0, \bar{t})$. By

Lemma 4, we know that $z = 0$. Since type $\bar{t}$ is of measure zero, its choice is inconsequential, but nevertheless we can still show that it's $S_0(\bar{t})$. If this is false, then, by Corollary 1, $Q(\bar{t}) > S_0(\bar{t})$. By Theorem 2, $S_0(\bar{t}) > x^*(\bar{t})$. Therefore, type t_0 must strictly prefer $(S_0(\bar{t}), \bar{t})$ to $(Q(\bar{t}), \bar{t})$. Take any sequence $t_k \uparrow \bar{t}$. By continuity of S_0, we know that $(S_0(t_k), t_k)$ converges to $(S_0(\bar{t}), \bar{t})$. But then type $\bar{t}$ would prefer to imitate type t_k for large k.

The final step in the proof is to show that these equilibria satisfy the D1 condition. Consider any equilibrium, and let x be some out-of-equilibrium action. For each type t, define $m_x(t)$ as the value of m that satisfies $U(x, m, t) = U(Q(t), B\left(\Phi_{Q(t)}\right), t)$ (if the left-hand side exceeds the right-hand side for all $m \geq 0$, then $m_x(t) = 0$; otherwise, existence and uniqueness of a solution to the equation is guaranteed because the left-hand side increases continuously and without bound in m). Let $M_x = \left\{t \in \left[0, \bar{t}\right] \mid m_x(t) \leq m_x(t') \forall t' \in \left[0, \bar{t}\right]\right\}$. The D1 criterion implies that Φ_x places probability only on the set M_x. Clearly, one can always find beliefs that satisfy this condition. The question is whether the beliefs would then induce some type to deviate from its equilibrium choice.

Begin with a separating equilibrium. Out-of-equilibrium actions consist of the interval $(S_0(\bar{t}), 1]$. Since no type t prefers $(S_0(\bar{t}), \bar{t})$ to its own equilibrium outcome, and since $S_0(\bar{t}) > S_0(t) > x^*(t)$, no type prefers (x, m) to its equilibrium choice for any $x > S_0(\bar{t})$ and any $m \leq \bar{t}$. Thus, the equilibrium is consistent with any inference Φ_x that places probability only on M_x.

Now consider a central pooling equilibrium. Out-of-equilibrium actions consist of the sets $(S_0(t_0), \frac{1}{2})$ and $(\frac{1}{2}, 1]$. Consider any action x in either of these intervals. We know that

$$F\left(1 - x, m_x(t_0)\right) - F\left(\frac{1}{2}, B\left(H_{t_0}\right)\right) = t_0 \left[G\left(0\right) - G\left(x - \frac{1}{2}\right)\right]$$

So, for any $t > t_0$,

$$F\left(1 - x, m_x(t_0)\right) - F\left(\frac{1}{2}, B\left(H_{t_0}\right)\right) < t \left[G\left(0\right) - G\left(x - \frac{1}{2}\right)\right]$$

This implies $m_x(t) > m_x(t_0)$, which in turn implies $t \notin M_x$. So, for any inference Φ_x that places probability only on M_x, $B(\Phi_x) \le t_0$. We will show that, for any such inference, no type would deviate to x.

Consider first $x > \frac{1}{2}$. For any such Φ_x, we have $B(H_{t_0}) > t_0 \ge B(\Phi_x)$, so every type t strictly prefers $\left(\frac{1}{2}, B(H_{t_0})\right)$ to $(x, B(\Phi_x))$. But since each type t weakly prefers $\left(Q(t), B\left(\Phi_{Q(t)}\right)\right)$ to $\left(\frac{1}{2}, B(H_{t_0})\right)$, no type has an incentive to deviate to x.

Now consider $x \in (S_0(t_0), \frac{1}{2})$. For $t \le t_0$, $(x, B(\Phi_x))$ is not as good as $(S_0(t_0), t_0)$ (because $x > S_0(t_0) \ge S_0(t) > x^*(t)$ and $B(\Phi_x) \le t_0$). Since type t's equilibrium payoff is at least as high as its payoff from choosing $(S_0(t_0), t_0)$, it has no incentive to select x. Moreover, since

$$F(1-x, B(\Phi_x)) - F\left(\frac{1}{2}, B(H_{t_0})\right) < t_0 \left[G(0) - G\left(x - \frac{1}{2}\right)\right]$$

we also have

$$F(1-x, B(\Phi_x)) - F\left(\frac{1}{2}, B(H_{t_0})\right) < t \left[G(0) - G\left(x - \frac{1}{2}\right)\right]$$

for $t > t_0$, which implies none of these types has an incentive to select x either. □

The following lemma is used in several of the remaining proofs, and requires Assumption B-2.

Lemma 5: Consider the CDFs J, K, and L, such that $J(t) = \lambda K(t) + (1-\lambda)L(t)$. If max *supp*$(L) \le B(J)$, then $B(J) \le B(K)$, where the second inequality is strict if the first is strict or if the support of L is nondegenerate.

Proof: Begin with the case where max *supp*$(L) < B(J)$. Clearly, $B(K) = B(J)$ would directly contradict Assumption B-2. Suppose on the contrary that $B(K) < B(J)$. Clearly, $\lambda < 1$. Define $J_\mu(t) = \mu K(t) + (1-\mu)L(t)$. Then $J_1 = K$ and $J_\lambda = J$. Since J_μ is continuous in μ (applying the weak topology), so is $B(J_\mu)$. Consequently, $\exists \mu' \in (\lambda, 1)$ such that

$$\max supp\,(L) < B(J_{\mu'}) < B(J)$$

Note that

$$\begin{aligned} J(t) &= \left(\frac{\lambda}{\mu'}\right)(\mu' K(t) + (1-\mu')L(t)) + \left(\frac{\mu'-\lambda}{\mu'}\right)L(t) \\ &= \gamma J_{\mu'}(t) + (1-\gamma)L(t) \end{aligned}$$

where $\gamma = \frac{\lambda}{\mu'} \in [0,1)$. By Assumption B-2, we know that $B(J) < B(J_\mu)$, a contradiction.

Now consider the case where max $supp(L) = B(J)$. Consider a sequence K^n converging to K in the weak topology where K^n is "higher" than K^{n+1} in the sense of first-order stochastic dominance. Let $J^n(t) = \lambda K^n(t) + (1-\lambda)L(t)$. Then, by Assumption B-1, we have $B(J^n) > B(J) = \max supp(L)$. By the preceding argument, $B(J^n) \le B(K^n)$ for all n. Since B is continuous, $B(J) \le B(K)$.

Now we consider the case where max $supp(L) = B(J)$ and the support of L is nondegenerate. We know from the previous argument that $B(K) \ge B(J)$. Suppose, contrary to the lemma, that $B(K) = B(J)$. Define L' as the CDF placing all probability on $B(J)$, and define $J'(t) = \lambda K(t) + (1-\lambda)L'(t)$. Since L' is "higher" than L in the sense of first-order stochastic dominance, J' is "higher" than J, so $B(J') > B(J)$. Since we also know that $B(K) = B(J) = \max supp(L')$, Assumption B-2 implies $B(J') \le B(K)$, a contradiction. □

Proof of Theorem 6: First, we claim that $S_0(t) < \widetilde{S}_0(t)$ for all t. Notice that

$$S_0'(0) = \frac{F_2(1,0)}{F_1(1,0)} < \frac{F_2(1,0) + \phi'(0)}{F_1(1,0)} = \widetilde{S}_0'(0)$$

It follows immediately that $S_0(t) < \widetilde{S}_0(t)$ for small t. Now we show that the same inequality holds for all t. Suppose the claim is false. Then, since the separating functions are continuous, the set $\left\{t > 0 \mid S_0(t) = \widetilde{S}_0(t)\right\}$ must be non-empty and compact. Let t' be the smallest element of this set. Letting $s^* \equiv S_0(t') = \widetilde{S}_0(t')$, we have

$$\begin{aligned} S_0'(t) &= \frac{F_2(1-s^*, t')}{F_1(1-s^*, t') - tG'(s^* - \frac{1}{2})} \\ &< \frac{F_2(1-s^*, t') + \phi(t)}{F_1(1-s^*, t') - tG'(s^* - \frac{1}{2})} = \widetilde{S}_0'(t) \end{aligned}$$

Moreover, since the slopes of the separating functions vary continuously with t, we must also have $S_0'(t) < \widetilde{S}_0'(t)$ for all t within some ε-neighborhood of t' (with $\varepsilon < t'$). But then

$$\begin{aligned}\widetilde{S}_0(t') &= \widetilde{S}_0(t'-\varepsilon) + \int_{t'-\varepsilon}^{t'} \widetilde{S}'(t)dt \\ &> S_0(t'-\varepsilon) + \int_{t'-\varepsilon}^{t'} S'(t)dt = S_0(t'),\end{aligned}$$

a contradiction.

It is trivial to check that $\widetilde{t}^*_{0,0} < t^*_{0,0}$. Therefore, if a central pooling equilibrium exists for U (that is, if $\bar{t} > t^*_{0,0}$), a central pooling equilibrium also exists for $\widetilde{U}$ (that is, $\bar{t} > \widetilde{t}^*_{0,0}$).

Next we define the function $\widetilde{\psi}(t)$ analogously to our definition of $\psi(t)$ in the proof of Theorem 4:

$$\widetilde{U}(\widetilde{S}_0(t), t, t) = \widetilde{U}\left(\frac{1}{2}, \widetilde{\psi}(t), t\right) \tag{13}$$

Notice that we can rewrite this definition as

$$U(\widetilde{S}_0(t), t, t) = U\left(\frac{1}{2}, \widetilde{\psi}(t), t\right) + \phi(\widetilde{\psi}(t)) - \phi(t)$$

Because $\widetilde{S}_0(t) \geq S_0(t) \geq x^*(t)$ and $\psi(t) > t$, we have

$$\begin{aligned}U(\widetilde{S}_0(t), t, t) &\leq U(S_0(t), t, t) = U\left(\frac{1}{2}, \psi(t), t\right) \\ &< U\left(\frac{1}{2}, \psi(t), t\right) + \phi(\psi(t)) - \phi(t)\end{aligned}$$

Since $U\left(\frac{1}{2}, m, t\right) + \phi(m)$ is increasing in m, the value of $\widetilde{\psi}(t)$ that solves (13) must be less than $\psi(t)$; thus, $\widetilde{\psi}(t) < \psi(t)$.

Let's focus on the cases for which there is a central pooling equilibrium with U (that is, $\bar{t} > t^*_{0,0}$). As we explained in the proof of Theorem 4, if $B(H) - \psi(0) \geq 0$, then there is a unique equilibrium in which $t_0 = 0$. Because $\widetilde{\psi}(t_0) < \psi(t_0)$, we have $B(H) - \widetilde{\psi}(0) > 0$, so there is also a unique equilibrium with $\widetilde{U}$ in which $\widetilde{t}_0 = 0$. We also explained in the proof of Theorem 4 that, if $B(H) - \psi(0) < 0$, then there is a unique equilibrium where t_0 is given by the solution to the equation $B(H_{t_0}) - \psi(t_0) = 0$. Note that, for that same value of

t_0, $B(H_{t_0}) - \widetilde{\psi}(t_0) > 0$. Therefore, since $\widetilde{\psi}(t)$ and $B(H_t)$ are, respectively, decreasing and increasing in t, either there is some $\widetilde{t}_0 < t_0$ that solves $B(H_{\widetilde{t}_0}) - \widetilde{\psi}(\widetilde{t}_0) = 0$, or $B(H) - \widetilde{\psi}(0) > 0$, in which case $\widetilde{t}_0 = 0$. In either case, we have $\widetilde{t}_0 < t_0$, as the theorem asserts. $\square$

Proof of Theorem 7: We begin by providing a formula for $\widehat{H}_{t_0}$. Applying Bayes' Law, the probability that D's type is less than or equal to t conditional on D choosing zero is

$$\widehat{H}_{t_0}(t) = \left(\frac{p}{p + (1-p)H(t_0)}\right) H(t) + \left(\frac{1-p}{p + (1-p)H(t_0)}\right) H\left(\min\{t_0, t\}\right)$$

Now we establish the result through a series of lemma.

Lemma 6: Equation (2) always has a unique solution, $t_0^* \in (0, \bar{t})$.

Proof: First rewrite (2) as

$$F\left(1, B\left(\widehat{H}_{t_0}\right)\right) + t_0 G\left(-\frac{1}{2}\right) = F\left(1 - x^*(t_0), t_0\right) + t_0 G\left(x^*(t_0) - \frac{1}{2}\right)$$

Define the function $\xi(t)$ as the solution to

$$F\left(1, \xi(t)\right) + tG\left(-\frac{1}{2}\right) = F\left(1 - x^*(t), t\right) + tG\left(x^*(t) - \frac{1}{2}\right) \tag{14}$$

A unique solution always exists because (1) for $\xi = t$, by the definition of $x^*(t)$, the left hand side does not exceed the right, (2) for sufficiently large ξ, the left hand side does exceed the right (F is unbounded in ξ), and (3) F is strictly increasing in ξ. Indeed, from these observations, it follows immediately that $\xi(t) \geq t$, with strict inequality if $x^*(t) > 0$. Note that we can rewrite a solution to (2) as

$$\xi(t_0) = B\left(\widehat{H}_{t_0}\right) \tag{15}$$

We proceed in a series of steps.

Step 1: There exists a solution to (15) with $t_0 \in (0, \bar{t})$. Since $x^*(0) = 0$, equation (14) implies that $\xi(0) = 0$. Since $\widehat{H}_0 = H$ and $B(H) > 0$, we have $\xi(0) < B\left(\widehat{H}_0\right)$. Further, since $\xi(\bar{t}) \geq \bar{t} > B(H)$, and since $\widehat{H}_{\bar{t}} = H$, we have $\xi(\bar{t}) > B\left(\widehat{H}_{\bar{t}}\right)$. By continuity, there must exist at least one solution between these extremes. The rest of the proof demonstrates uniqueness.

Step 2: $\xi(t)$ is strictly monotonically increasing. For $t \neq t^*$ (defined in Theorem 1), implicitly differentiating (14) (and applying the envelope theorem for $t > t^*$) we obtain

$$\xi'(t) = \frac{F_2\left(1 - x^*(t), t\right) + \left[G\left(x^*(t) - \frac{1}{2}\right) - G\left(-\frac{1}{2}\right)\right]}{F_2\left(1, \xi(t)\right)}$$

Since $F_2\left(1 - x^*(t), t\right) > 0$, $F_2\left(1, \xi(t)\right) > 0$, and $G\left(x^*(t) - \frac{1}{2}\right) - G\left(-\frac{1}{2}\right) \geq 0$, the entire term is strictly positive.

Step 3: Consider some t', t'' with $t' > t''$. Then

$$\widehat{H}_{t'}(t) = \lambda \widehat{H}_{t''}(t) + (1 - \lambda) L(t) \tag{16}$$

where

$$\lambda = \frac{p + (1-p)H(t'')}{p + (1-p)H(t')} \in (0, 1)$$

and $L(t)$ is a CDF given by

$$L(t) = \frac{H\left(\min\{t, t'\}\right) - H\left(\min\{t, t''\}\right)}{H(t') - H(t'')} \tag{17}$$

Demonstrating this property is just a matter of algebra. Notice that max $supp(L) = t'$ (i.e., $L(t') = 1$).

Step 4: Uniqueness. Suppose, contrary to the lemma, that there are two solutions, t' and t'', with $t' > t''$. We know that $t' \leq \xi(t') = B\left(\widehat{H}_{t'}\right)$. From this, we know that max $supp(L) \leq B\left(\widehat{H}_{t'}\right)$ (where L is defined in (17)). Noting (16) and applying Lemma 5, we have $B\left(\widehat{H}_{t'}\right) \leq B\left(\widehat{H}_{t''}\right)$. From step 2 and the fact that $t' > t''$, we know that $\xi(t') > \xi(t'')$. Putting these facts together, we have

$$\xi(t'') < \xi(t') = B\left(\widehat{H}_{t'}\right) \leq B\left(\widehat{H}_{t''}\right),$$

which contradicts the supposition that t'' is a solution. □

Henceforth in this proof, wherever we write t_0^*, we mean the solution to (2).

Lemma 7: Suppose (5) holds. Then no Case A equilibrium exists.

Proof: First consider Case A-1. Since $t_1 > t_0^*$, we know that

$$U\left(x^*(t_0^*), t_0^*, t_0^*\right) \leq U\left(\frac{1}{2}, B\left(H_{t_0^*}\right), t_0^*\right) < U\left(\frac{1}{2}, B\left(H_{t_1}\right), t_0^*\right)$$

But then t_0^* will choose $x = \frac{1}{2}$ rather than $x = Q(t_0^*) = x^*(t_0^*)$, a contradiction.

Now consider Case A-2. Let x_0 solve $\max_x U\left(x, \bar{t}, t_0^*\right)$. We claim that $x_0 \leq x^*(\bar{t})$. Suppose on the contrary that $x_0 > x^*(\bar{t})$. We know that

$$t_0^*\left(G\left(x_0 - \frac{1}{2}\right) - G\left(x^*(\bar{t}) - \frac{1}{2}\right)\right) \geq F\left(1 - x^*(\bar{t}), \bar{t}\right) - F\left(1 - x_0, \bar{t}\right)$$

But then

$$\bar{t}\left(G\left(x_0 - \frac{1}{2}\right) - G\left(x^*(\bar{t}) - \frac{1}{2}\right)\right) \geq F\left(1 - x^*(\bar{t}), \bar{t}\right) - F\left(1 - x_0, \bar{t}\right)$$

which contradicts the fact that $x^*(\bar{t})$ solves $\max_x U\left(x, \bar{t}, \bar{t}\right)$.

Since $x_0 \leq x^*(\bar{t}) < S^{t_0^*}(\bar{t}) \leq \frac{1}{2}$, we know that $U\left(\frac{1}{2}, \bar{t}, t_0^*\right) \leq U\left(S^{t_0^*}(\bar{t}), \bar{t}, t_0^*\right)$. This implies

$$\begin{aligned} U\left(x^*(t_0^*), t_0^*, t_0^*\right) &\leq U\left(\frac{1}{2}, B\left(H_{t_0^*}\right), t_0^*\right) \\ &< U\left(\frac{1}{2}, \bar{t}, t_0^*\right) \\ &\leq U\left(S^{t_0^*}(\bar{t}), \bar{t}, t_0^*\right) \end{aligned}$$

But then t_0^* will choose $x = Q(\bar{t}) = S^{t_0^*}(\bar{t})$ rather than $x = Q(t_0^*) = x^*(t_0^*)$, a contradiction. □

Lemma 8: Suppose (6) holds. Then there exists a unique Case A equilibrium. If $S^{t_0^*}(\bar{t}) > \frac{1}{2}$, it is a Case A-1 equilibrium, and $S^{t_0^*}(t_1) < \frac{1}{2}$. If $S^{t_0^*}(\bar{t}) \leq \frac{1}{2}$, it is a case A-2 equilibrium.

Proof: We know from Lemma 6 that the value of t_0^* is uniquely determined. The analysis of signaling equilibria for types $[t_0^*, \bar{t}]$ mirrors the analysis for all types $[0, \bar{t}]$ in Section 4. Accordingly, the proof of this lemma involves essentially the same arguments as Theorems 3 (for Case A-2 equilibria) and 4 (for Case A-1 equilibria). □

Lemma 9: When (6) holds, a Case B equilibrium does not exist. When (5) holds, there exists a unique Case B equilibrium.

Proof: We prove the lemma through a series of steps.

Step 1: For any t_0 satisfying (4), $t_0 \in (0, t_0^*]$. Moreover, for any $t \in (0, t_0^*]$, $U\left(0, B\left(\widehat{H}_t\right), t\right) \geq U\left(x^*(t), t, t\right)$.

The proof of Lemma 6 establishes that $U\left(0, B\left(\widehat{H}_t\right), t\right) - U\left(x^*(t), t, t\right)$ exceeds zero for $t = 0$, is less than zero for $t = \bar{t}$, and equals zero at a unique value $t = t_0^*$. Since the expression is continuous in t, it must be negative for $t \geq t_0^*$ (otherwise t_0^* would not be unique). Consequently, we must have $t_0 \leq t_0^*$. Since $U\left(0, B\left(H\right), 0\right) > U\left(\frac{1}{2}, B\left(H\right), 0\right)$, we know $t_0 \neq 0$.

For the remainder of the proof, we define the function $\zeta(t)$ as follows: (1) if $U\left(0, 0, t\right) \geq U\left(\frac{1}{2}, B\left(H_t\right), t\right)$, then $\zeta(t) = 0$; (2) if $U\left(0, 0, t\right) < U\left(\frac{1}{2}, B\left(H_t\right), t\right)$, then $\zeta(t)$ solves

$$U\left(0, \zeta(t), t\right) = U\left(\frac{1}{2}, B\left(H_t\right), t\right). \tag{18}$$

Existence, uniqueness, and continuity of $\zeta(t)$ are all easy to verify. Moreover, the equality in (4) is equivalent to the statement that

$$\zeta(t_0) = B\left(\widehat{H}_{t_0}\right) \tag{19}$$

(since $B\left(\widehat{H}_{t_0}\right) > 0$, there is never a solution with $\zeta(t_0) = 0$, which is the only circumstance where the equality in (4) would not follow immediately from (19)). From step 1, we know that t_0 satisfies (4) iff $t_0 \in (0, t_0^*]$ and (19) holds.

Step 2: There exists $\widehat{t}$ such that $\zeta(t) = 0$ for $t < \widehat{t}$, $\zeta(t)$ is strictly increasing in t for $t \geq \widehat{t}$.

We can rewrite the equation (18) as

$$F\left(1, \zeta(t)\right) = t\left(G\left(0\right) - G\left(-\frac{1}{2}\right)\right) + F\left(\frac{1}{2}, B(H_t)\right)$$

Note that the right-hand side of this expression is strictly increasing in t, and the left-hand side is strictly increasing in ς. Therefore, if $\zeta(t) > 0$ and $t' > t$, we must have $\zeta(t') > \zeta(t)$.

Thus, if $\zeta(0) > 0$, we can trivially take $\widehat{t} = 0$; if $\zeta(0) = 0$, we can take $\widehat{t}$ to be the maximal $t \in [0, \overline{t}]$ for which $\zeta(t) = 0$.

Step 3: $B\left(\widehat{H}_t\right)$ is weakly decreasing in t for $t \in (0, t_0^*]$.

Consider any two values, t', $t'' \leq t_0^*$, with $t' > t''$. We know that $t' \leq \xi(t') \leq B\left(\widehat{H}_{t'}\right)$ (this follows from the proof of Lemma 6, given $t' < t_0^*$). >From this, we know that max $supp(L) \leq B\left(\widehat{H}_{t'}\right)$ (where L is defined in (17) from step 3 of the proof of lemma 6). Noting (16) and applying Lemma 5, we have $B\left(\widehat{H}_{t'}\right) \leq B\left(\widehat{H}_{t''}\right)$.

Step 4: When (6) holds, a Case B equilibrium does not exist.

In this case, we have

$$U\left(0, B\left(\widehat{H}_{t_0^*}\right), t_0^*\right) = U\left(x^*(t_0^*), t_0^*, t_0^*\right) > U\left(\frac{1}{2}, B\left(H_{t_0^*}\right), t_0^*\right)$$

But this means $\zeta(t_0^*) < B\left(\widehat{H}_{t_0^*}\right)$. From steps 2 and 3, we know $\zeta(t) < B\left(\widehat{H}_t\right)$ for all $t < t_0^*$. Consequently, there exists no t_0 satisfying (4).

Step 5: When (5) holds, there exists a unique Case B equilibrium, with $t_0 \in (\widehat{t}, t_0^*]$.

In this case, we have

$$U\left(0, B\left(\widehat{H}_{t_0^*}\right), t_0^*\right) = U\left(x^*(t_0^*), t_0^*, t_0^*\right) \leq U\left(\frac{1}{2}, B\left(H_{t_0^*}\right), t_0^*\right)$$

This implies that $\zeta(t_0^*) \geq B\left(\widehat{H}_{t_0^*}\right) > 0$, from which it also follows that $\widehat{t} < t_0^*$.

We claim that $\zeta(t) < B\left(\widehat{H}_t\right)$ for all $t \leq \widehat{t}$. If $\widehat{t} > 0$, the claim follows immediately from the fact that $\zeta(t) = 0 < B\left(\widehat{H}_t\right)$ for all $t \leq \widehat{t}$. If $\widehat{t} = 0$, we have $U(0, \zeta(0), 0) = U\left(\frac{1}{2}, B(H_0), 0\right)$, from which it follows that $\zeta(0) < B(H_0)$. But $B(H_0) = B(H) = B\left(\widehat{H}_0\right)$, so we also have $\zeta(0) < B\left(\widehat{H}_0\right)$ as desired.

By continuity, a solution to $\zeta(t) = B\left(\widehat{H}_t\right)$ exists on $(\widehat{t}, t_0^*]$. Since $\zeta(t)$ is strictly increasing in t on $[\widehat{t}, t_0^*]$, and since $B\left(\widehat{H}_t\right)$ is weakly decreasing, the solution is unique. $\square$

Lemmas 7, 8, and 9 establish the Theorem. $\square$

Proof of Theorem 8: Most of the arguments in Theorem 5 apply directly. In particular, the arguments establish, without modification, that (1) no type chooses $x > \frac{1}{2}$, (2) choices are weakly monotonic in type, (3) there is no pool at any action other than 0 and $\frac{1}{2}$,[22] and (4) if $Q(t) < x$ where x is an action not chosen in equilibrium, the social image associated with x is no less than t.

We argue first that there is positive mass at $x = 0$. Suppose not. Then the social image associated with $x = 0$ is $B(H)$. Type $t = 0$ chooses some $x > 0$. Since choices are monotonic in type, the associated social image cannot exceed $B(H)$ (and only achieves this upper bound if all types make the same choice). But type $t = 0$ would plainly prefer to pick $x = 0$ and receive social image $B(H)$.

Next we argue that some types do not select $x = 0$. Suppose this is false. Then the social image associated with $x = 0$ is $B(H)$, and the social image associated with all $x > 0$ is $\bar{t}$ (see property (4) above). In that case, all types could beneficially deviate to some x slightly greater than zero.

From these arguments, we know that an equilibrium must take one of the following forms: (a) all mass split between zero and $\frac{1}{2}$, (b) mass at zero, with an atomless distribution of choices between zero and $\frac{1}{2}$, and no mass at $\frac{1}{2}$, or (c) mass at zero and $\frac{1}{2}$, with an atomless distribution of choices between these extremes.

Consider possibility (a). Since actions are monotonic in types, there exists $t_0 \in (0, \bar{t})$ such that $t \in [0, t_0]$ choose $x = 0$ and $t \in (t_0, \bar{t}]$ choose $x = \frac{1}{2}$ (recall that our convention is to break indifference in favor of the lower group). Clearly, t_0 must be indifferent between $(x, m) = \left(0, B\left(\widehat{H}_{t_0}\right)\right)$ and $(x, m) = \left(\frac{1}{2}, B\left(H_{t_0}\right)\right)$, or some type close to t_0 would deviate to the other pool. Thus, the equality in (4) must be satisfied. In addition, the social image associated with all $x \in \left(0, \frac{1}{2}\right)$ must be at least t_0, so the inequality in (4) must be satisfied. Thus, possibility (a) must be a Case B equilibrium.

Consider possibility (b). Since actions are monotonic in types, there exists $t_0 \in (0, \bar{t})$ such that $t \in [0, t_0]$ choose $x = 0$ and $t \in (t_0, \bar{t}]$ choose separating actions. We claim that, for

[22] The argument ruling out pools at actions other than $\frac{1}{2}$ relied on the fact that, for any pool, the social image of the pool is less than the highest type in the pool. With exogenous selection of zero, this argument does not rule out a pool at zero, since the social image of the pool may exceed (or equal) the highest type in the pool.

all $t \in (t_0, \overline{t}]$, $Q(t) \geq x^*(t)$. Suppose on the contrary that $Q(t') < x^*(t')$ for some $t' \in (t_0, \overline{t}]$. Let m^* denote the social image associated with choice $x^*(t')$. We know that $m^* \geq t'$. But then

$$U(Q(t'), t', t') < U(x^*(t'), t', t') \leq U(x^*(t'), m^*, t'),$$

which implies that t' could deviate beneficially to $x^*(t')$, a contradiction.

Next we claim that $\lim_{t \downarrow t_0} Q(t) = x^*(t_0)$. If this is not the case then, by the previous claim, $\lim_{t \downarrow t_0} Q(t) > x^*(t_0)$. Let m^* be the social image associated with $x^*(t_0)$; we know that $m^* \geq t_0$. Note that

$$\begin{aligned} \lim_{t \downarrow t_0} U(Q(t), t, t) &= U\left(\lim_{t \downarrow t_0} Q(t), t_0, t_0\right) \\ &< U(x^*(t_0), t_0, t_0) \\ &\leq U(x^*(t_0), m^*, t_0) \end{aligned}$$

But then, for $t' > t_0$ with $t' - t_0$ small, $U(Q(t'), t', t') < U(x^*(t_0), m^*, t')$. This implies that t' could beneficially deviate to $x^*(t_0)$.

From the previous two claims, we know that, for $t > t_0$, $Q(t)$ must be the separating function initialized at $(t_0, x^*(t_0))$, that is, $S^{t_0}(t)$.

Clearly, t_0 must be indifferent between $(x, m) = \left(0, B\left(\widehat{H}_{t_0}\right)\right)$ and $(x, m) = (x^*(t_0), t_0)$, or some type in a small neighborhood of t_0 could beneficially deviate. Thus, (2) must be satisfied, which implies that possibility (b) must be a Case A-2 equilibrium.

Consider possibility (c). Since actions are monotonic in types, there exists $t_0, t_1 \in (0, \overline{t})$ with $t_0 > t_1$ such that $t \in [0, t_0]$ choose $x = 0$, $t \in (t_0, t_1]$ choose separating actions, and $t \in (t_1, \overline{t}]$ choose $x = \frac{1}{2}$. The arguments provided for possibility (b) implies that, for $t \in (t_0, t_1]$, $Q(t) = S^{t_0}(t)$, and that (2) holds. In addition, t_1 must be indifferent between $(x, m) = (S^{t_0}(t_1), t_1)$ and $(x, m) = \left(\frac{1}{2}, B(H_{t_1})\right)$, or some type in a small neighborhood of t_0 could beneficially deviate. This implies that (3) holds. Thus, possibility (c) must be a Case A-1 equilibrium.

To complete the proof, we need only demonstrate that Case A and Case B equilibria satisfy the D1 criterion, given an appropriate choice of beliefs. The argument is essentially the same as that given in the proof of Theorem 5 (that is, inferences assign unused actions to the highest type taking a lower action). □

Proof of Theorem 9: We begin by introducing some notation. To reflect the dependence of t_0^* on p, we will use the notation $t_0^*(p)$. Define $t_1^*(p)$ as follows:

$$t_1^*(p) = \begin{cases} t_0^*(p) \text{ when (5) holds} \\ \bar{t} \text{ when (6) holds and } S^{t_0^*(p)}(\bar{t}) \leq \frac{1}{2} \\ \text{the solution to (3) otherwise} \end{cases} \tag{20}$$

We note that $t_1^*(p)$ is well-defined for all p. The three cases listed above correspond to the circumstances in which there is, respectively, a Case B equilibrium, a Case A-2 equilibrium and a Case A-1 equilibrium. These cases are mutually exclusive and exhaustive, and in the third case we know a solution to (3) exists. Finally, define $\widehat{t}_0(p)$ as follows:

$$\widehat{t}_0(p) = \begin{cases} \text{the solution to the equality in (4) when (5) holds} \\ t_0^*(p) \text{ when (6) holds} \end{cases}$$

We note that $\widehat{t}_0(p)$ is well-defined for all p. The two cases listed above correspond to the circumstances in which there is, respectively, a Case B equilibrium, and a Case A equilibrium. These cases are mutually exclusive and exhaustive, and in the first case we know that a solution to (4) exists.

We have constructed these functions so that, given p, in an equilibrium types $t \in [0, \widehat{t}_0(p)]$ choose $x = 0$, and types $t \in (t_1^*(p), \bar{t}]$ choose $x = \frac{1}{2}$ (regardless of whether Case A-1, Case A-2, or Case B prevails). The proof of the theorem involves showing that both of these functions are continuous and strictly increasing in p.

Step 1: $\widehat{t}_0^*(p)$ and $t_1^*(p)$ are both continuous in p.

Begin with $\widehat{t}_0^*(p)$. First we claim that $t_0^*(p)$ is continuous. This follows from uniqueness and continuity of the functions in (2). Next we claim that the solution to (4) is continuous in p for p such that (5) holds. This follows from uniqueness and continuity of the functions in the equality of (4). Finally, we argue that, if (5) holds with equality for some p (the boundary case between (5) and (6)), then $\widehat{t}_0(p) = t_0^*(p)$. In this case, we have

$$U\left(0, B\left(\widehat{H}_{t_0^*(p)}\right) t_0^*(p)\right) = U\left(x^*\left(t_0^*(p)\right), t_0^*(p), t_0^*(p)\right) = U\left(\frac{1}{2}, B\left(H_{t_0^*(p)}\right) t_0^*(p)\right)$$

(where the first equality follows from the definition of $t_0^*(p)$, and the second follows from the fact that (5) holds with equality). But then $t_0 = t_0^*(p)$ satisfies (4), so $\widehat{t}_0(p) = t_0^*(p)$.

Now we turn to $t_1^*(p)$. Since (1) $t_0^*(p)$ is continuous in p, (2) $S^{t_0}(t)$ is continuous for all t_0 and t, and (3) the solution to (3), when it exists, is unique, we know that the solution to (3) is continuous in p (when it exists).

Next we argue that, if (5) holds with equality for some p, then $S^{t_0^*(p)}(\bar{t}) > \frac{1}{2}$ (so that this is the boundary case between the first and third possibilities in (20)) and $t_0^*(p)$ solves (3). The latter statement follows immediately from the fact that $S^{t_0^*(p)}\left(t_0^*(p)\right) = x^*(t_0^*(p))$. Suppose contrary to the first statement that $S^{t_0^*(p)}(\bar{t}) \leq \frac{1}{2}$. We know that

$$\begin{aligned} U\left(S^{t_0^*(p)}(\bar{t}), \bar{t}, t_0^*(p)\right) &\geq U\left(\frac{1}{2}, \bar{t}, t_0^*(p)\right) \\ &> U\left(\frac{1}{2}, B\left(H_{t_0^*(p)}\right), t_0^*(p)\right) \\ &= U\left(x^*\left(t_0^*(p)\right), t_0^*(p), t_0^*(p)\right), \end{aligned}$$

which contradicts the fact that $S^{t_0^*(p)}$ is a separating function ($t_0^*(p)$ would imitate $\bar{t}$).

Finally, we note that, if (6) holds and $S^{t_0^*(p)}(\bar{t}) = \frac{1}{2}$ (the boundary case between the second and third possibilities in (20)), then $\bar{t}$ solves (3) (this follows immediately from inspection of (3) given that $B\left(H_{\bar{t}}\right) = \bar{t}$).

Step 2: $t_0^*(p)$ is strictly increasing in p.

Recall that $t^*(p)$ is the solution to (15). From (14), we know that $\xi(t)$ is independent of p. We have already shown that $\xi(t)$ is strictly increasing in t (step 2 in the proof of Lemma 6). We have also shown that, for $\tau \leq t_0^*(p)$, $B\left(\widehat{H}_\tau^p\right)$ is weakly decreasing in τ (step 3 in the proof of Lemma 9). We claim that, if $p' > p''$, then $B\left(\widehat{H}_\tau^{p'}\right) > B\left(\widehat{H}_\tau^{p''}\right)$ for $\tau \leq t_0^*(p'')$. It follows directly from this claim that $B\left(\widehat{H}_\tau^{p'}\right) > \xi(\tau)$ for $\tau \leq t_0^*(p'')$, so $t_0^*(p') > t_0^*(p'')$, as desired.

We now prove the claim. Define

$$\widehat{H}_\tau^p(t) = \left(\frac{p}{p+(1-p)H(\tau)}\right) H(t) + \left(\frac{1-p}{p+(1-p)H(\tau)}\right) H\left(\min\{\tau, t\}\right)$$

Consider p', $p'' \in [0,1]$ with $p' > p''$. Note that

$$\widehat{H}_\tau^{p''}(t) = \lambda \widehat{H}_\tau^{p'}(t) + (1-\lambda)L(t) \tag{21}$$

where

$$\lambda = \left(\frac{p''}{p'}\right)\left(\frac{p'+(1-p')H(\tau)}{p''+(1-p'')H(\tau)}\right) \in (0,1)$$

and $L(t)$ is a CDF given by

$$L(t) = \frac{H\left(\min\{\tau, t\}\right)}{H(\tau)}$$

For $\tau \leq t_0^*(p'')$, max $supp(L) = \tau \leq \xi(\tau) \leq B\left(\widehat{H}_\tau^{p''}\right)$. Since the support of L is nondegenerate, $B\left(\widehat{H}_\tau^{p'}\right) > B\left(\widehat{H}_\tau^{p''}\right)$ by Lemma 5.

Step 3: Consider p', p'' with $p' > p''$, and suppose that a Case B equilibrium exists for p' and p''. Then $\widehat{t}_0(p') > \widehat{t}_0(p'')$.

Recall that the equality in (4) is equivalent to (19), and that $\widehat{t}_0(p'') \leq t_0^*(p'')$. From (18), we know that $\zeta(t)$ is independent of p. We have already shown that $\zeta(t)$ is weakly increasing in t, and strictly increasing when strictly positive (step 2 in the proof of Lemma 9). We have also shown that, for $\tau \leq t_0^*(p)$, $B\left(\widehat{H}_\tau^{p}\right)$ is weakly decreasing in τ (step 3 in the proof of Lemma 9). In the previous step, we demonstrated that, if $p' > p''$, then $B\left(\widehat{H}_\tau^{p'}\right) > B\left(\widehat{H}_\tau^{p''}\right)$ for $\tau \leq t_0^*(p'')$. It follows directly from this claim that $B\left(\widehat{H}_\tau^{p'}\right) > \zeta(\tau)$ for $\tau \leq \widehat{t}_0(p'')$, so $\widehat{t}_0(p') > \widehat{t}_0(p'')$, as desired.

Step 4: Consider p', p'' with $p' > p''$, and suppose that a Case A-1 equilibrium exists for p' and p''. Then $t_1^*(p') > t_1^*(p'')$.

We know from step 2 that $t_0^*(p') > t_0^*(p'')$. Since $S^{t_0^*(p'')}\left(t_0^*(p')\right) > x^*\left(t_0^*(p')\right) = S^{t_0^*(p')}\left(t_0^*(p')\right)$, we know that $S^{t_0^*(p')}(t) < S^{t_0^*(p'')}(t)$ for all $t > t_0^*(p')$.

Analogously to the proof of Theorem 4, we define $\psi^p(t)$ as the solution to

$$U\left(S^{t_0^*(p)}(t), t, t\right) = U\left(\frac{1}{2}, \psi^p(t), t\right) \tag{22}$$

We can rewrite the solution for $t_1^*(p)$ (when a Case A-1 equilibrium exists) as $\psi^p(t) = B(H_t)$. Arguing as in the proof of Theorem 4, one can show that $\psi^p(t)$ is decreasing and continuous in t, while $B(H_t)$ is increasing and continuous in t. Moreover, from the argument in the preceding paragraph, we know that an increase in p increases the left-hand side of (22), which means that it strictly increases $\psi^p(t)$. Thus, the value of t satisfying $\psi^p(t) = B(H_t)$ must rise.

All but the final sentence of the theorem follows from Steps 1-4. First consider $\widehat{t}_0(p)$. Step 2 shows that it increases monotonically when Case A prevails, step 3 shows that it increases monotonically when Case B prevails, and step 1 shows that it changes continuously as we move from one case to the other. Next consider $t_1^*(p)$. It is constant when Case A-2 prevails, step 2 shows that it increases monotonically when Case B prevails, step 4 shows that it increases nonmonotonically when Case A-1 prevails, and step 1 shows that it changes continuously as we move from one case to the other.

Finally, we argue that, as p goes to zero, t_0 goes to zero, so the measure of types choosing $x = 0$ converges to zero. Note that both (2) and (4) require $B\left(\widehat{H}_{t_0}\right) \geq t_0$. But, fixing any $t_0 > 0$, as p goes to zero, $B\left(\widehat{H}_{t_0}\right)$ falls below t_0. To satisfy the requirement, t_0 must converge to zero. $\square$