

Fair Shares: Accountability and Cognitive Dissonance in Allocation Decisions

By JAMES KONOW*

Everyone has observed people invoking fairness arguments in defense of their opinions or actions, and it is not uncommon for such arguments to be wielded on both sides of an issue about which views conflict. For example, during a televised debate Representative Charles B. Rangel said, "I think [Affirmative Action] has to involve a search for fairness," whereas commentator Avi Nelson opined that "you promote more unfairness than fairness when you depart from the basic criterion, which is that individuals should be treated as individuals" (Annenberg/CPB Collection, 1984). On the 1986 tax reform legislation, Senator Robert Packwood stated: "Taxes are about more than money and they're about more than economics. They're about fairness, and this bill is fair," whereas Senator Carl M. Levin argued: "For our economy, this is the wrong bill at the wrong time ... making deficit reduction more difficult and less fair" (*Los Angeles Times*, September 28, 1986 pp. A1, A8).

Such cases contribute to the frequent conclusion that justice is merely a ploy, a vacuous concept used opportunistically by self-interested and self-serving agents. If fairness arguments were sheer subterfuge, however, it would be difficult to account for their use at all, let alone their frequent use or earnest consideration by others. That they have at least occasional impact on outcomes may be inferred from the very fact that they are advanced. Those

who take justice seriously also claim more direct evidence of its effects on legal proceedings, government regulatory and taxation policies, wage and benefit structures in the workplace, results of bargaining experiments in the laboratory, and even pricing policies in the market (e.g., Daniel Kahneman et al., 1986b; R. Mark Isaac et al., 1991; H. Peyton Young, 1994; Linda Babcock et al., 1995).

On the other hand, the ostensibly self-serving manipulation of fairness arguments also presents a problem for proponents of justice and defenders of its importance in social interaction. If people value equity so highly that they make personal sacrifices for its sake, how is it that they also apparently distort it for their own selfish ends? Although most studies of bargaining behavior suggest nontrivial deviations from narrow self-interest (e.g., Werner Güth et al., 1982; Alvin E. Roth et al., 1991), some of these same studies reject the "fairness hypothesis" that fairness alone accounts for this behavior (e.g., Gary E. Bolton, 1991; Robert Forsythe et al., 1994).

This paper proposes and tests a theory of decision making that attempts to elucidate the roles of fairness, self-interest, and self-deception in the allocation of economic rewards. Fairness is treated as a genuine value, but there also exists an incentive and a potential for changing beliefs about it. This suggests an explanation for why arguments about it are advanced and considered, but sometimes conflict, as in the affirmative action and taxation examples cited earlier. In addition to the standard "material utility" from the agent's own allocation of a good, the theory integrates concepts of fairness and cognitive dissonance into the objective function.¹ When the decision maker has

* Department of Economics, Loyola Marymount University, 7900 Loyola Boulevard, Los Angeles, CA 90045. The author gratefully acknowledges the helpful comments and suggestions of two anonymous referees of this *Review* and of Reinhard Selten, Werner Güth, Gabriel Fuentes, Gary Biglaiser, and of seminar participants at the University of Bonn, the University of Arizona, Humboldt University, and the 1997 Economic Science Association meetings. I also thank Tim Cason, John Conlisk, Joseph Earley, Zaki Eusufzai, Gabriel Fuentes, and Petra Konow for their advice or assistance conducting the experiments, and the LMU Research Committee and College of Liberal Arts for financial support. Any remaining errors are, of course, my own.

¹ Material utility is what is meant when (narrow) self-interest or selfishness are mentioned in this paper. Of course, if fairness is assumed to be a goal of the agent, one may argue that it is in his or her self-interest to be, to some

a personal stake in, and influence over, the outcome of an allocation, material utility entices him or her to attempt to secure for him- or herself more than the fair amount. The conflicting desires for both self-interest and fairness create, in the terminology of social psychology, “cognitive dissonance,” that is, an unpleasant tension or disutility. As stated in Leon Festinger’s seminal work (1957) and many after it, cognitive dissonance theory proposes that the agent is motivated to reduce this tension and may, in this context, do so either by reducing self-interested behavior, or by engaging in self-deception (here by choosing to believe that it is fair to take more than the fair amount), or by some combination of the two.

Thus, agents are posited to trade off material utility with fairness and, in some cases, to reduce dissonance through self-deception. The framework for formally developing and testing this theory in the current paper involves variations of the so-called “dictator game.” In the standard version of this exercise subjects in one group (the dictators) allocate a fixed sum of money between themselves and anonymous counterparts in another group (the recipients), whereby the recipients have no recourse. Here standard and new versions of dictator experiments designed to test certain predictions of the theory provide corroborative evidence on both the fairness and cognitive dissonance components. In addition, these experiments yield a measure of the extent to which “unfair” behavior may be attributed to unadulterated self-interest or to self-deception and indicate a substantial role for the latter. This conclusion also suggests an explanation for the wide variation in the degree of generosity exhibited by dictators that has been observed in numerous runs of the standard dictator experiment elsewhere [see, for example, Kahneman et al. (1986a) and Elizabeth Hoffman et al. (1994)]. An implication of this paper is that such dispersion may be traced to the extent to which the procedures and instructional language of differ-

ent experiments facilitate the self-deceptive manipulation of “genuine” fairness concepts on the part of dictators.

Fairness as it emerges in real-life situations is often attributed to various and sometimes conflicting rules or principles [see, for example, Young (1994) and Edward E. Zajac (1995)]. My own take on fairness is that most peoples’ values may be accounted for by several fairly simple principles, any of which may dominate depending on the context. This view differs from the common belief that fairness is hopelessly amorphous or subjective, lacking any definite or enduring form. A closely related position is that justice is a complex or contingent concept, varying widely across context or, to whatever extent that justice does lend itself to generalization, depending on a multitude of potentially contradictory principles. This impression is certainly strengthened by the seemingly endless arguments advanced about fairness and the frustration experienced in attempting to find solutions to practical problems using fairness principles.

Agreement on *principles* of fairness, however, does not rule out substantial disparity in *claims* based on those principles. In addition, the difficulty, perhaps even impossibility, of simple *solutions* to injustice does not preclude the existence of simple *principles* of justice. In the author’s view, reported fairness—as it emerges in survey responses, laboratory decisions, and social interaction—is the product of true underlying values, which may be distorted by a variety of factors. People may weight competing justice principles differently or may perceive and evaluate the factors relevant to even a single principle differently [evidence of these points is summarized in Konow (1999)]. In addition, other goals, such as narrow self-interest, may dominate or bias the concern for equity and have an impact on behavior. Working in synergy, differing perceptions and competing goals may cause reported fairness to be distorted vis-à-vis the underlying values.

This paper focuses on a single rule of justice, the *accountability principle*, because, for these experiments, it seems to characterize most accurately both the undistorted views of fairness and the basis for self-deception. Roughly speaking, the accountability principle requires that a person’s fair allocation (e.g., of income) vary in

extent, fair. I shun this usage, however, because it makes self-interested behavior a tautology and the moral terminology void, a position that conflicts with most people’s view of self-interest and fairness as distinct, meaningful, and nonempty concepts.

proportion to the relevant variables that he can influence (e.g., work effort) but not according to those that he cannot reasonably influence (e.g., a physical handicap).

The specific experimental methodology for the empirical portion of this study was chosen to provide evidence on certain conclusions of the proposed theory of decision making. First, the accountability principle is tested with paid subjects whose decisions affect actual allocations, as opposed to the uncompensated respondents to attitude surveys in Konow (1996). Second, the experiments strive to establish whether and to what extent fairness, self-interest, and self-deception play a role in allocations. Third, more specific predictions of dissonance theory with respect to experimental variations are examined based on the outcomes of this and other studies.

This paper is organized as follows. Section I presents more formally the theory of decision making and the experimental design within which it is formulated. Section II describes the experimental procedures and Section III contains summaries and statistical analyses of the results. Section IV is a discussion, mostly analyzing the relationship of this study to others.

I. Theory and Design

As discussed previously, the theory of decision making integrates two other theories: one of fairness and the other of cognitive dissonance. They are presented separately below as applied to the dictator experiments designed to test them. In this section the essential aspects of the design are introduced parallel to the development of particular features of the model, whereas the details of the procedures are provided in the following section. All subjects were given equal show-up fees, and subjects in separate rooms were anonymously paired with one another. The experiments were conducted in two stages. In the first stage each pair prepared letters for mailing and, based on the number of letters produced, was given money credits to a joint account assigned to each pair. In the second stage, the money credits were allocated among the pair in any way seen fit by an arbitrarily chosen subject (the dictator). Various treatments were carried out that differed, as explained below, with respect to how credits were determined in the first phase and allocated in the second phase.

A. A Model of Fairness

Here a model is presented of how the money credits from preparing letters should be allocated among members of a pair according to the fairness theory introduced in Konow (1996). This theory is closest in genus to "equity theory," a view of distributive justice with origins in sociology and social psychology.² It attempts to characterize the fair allocation, or *entitlement*, of some economic variable, such as money, according to the values of people who have no personal stake in the outcome. Although material utility is set aside in this approach, fairness judgments are taken, in some measure, to be subjective and dependent on differing perceptions. Fairness is also a relative concept that involves comparisons among individuals, here among the two subjects producing the letters.

In evaluating the entitlement of subject i in this experiment, denoted η_i , a person may consider the perceived *output*, denoted q_i , or subject i 's production of the variable being allocated, here the amount of money credited to the joint account that is attributed to the subject. Also relevant is the perceived *input* x_i , or a measure of the person's contribution to the output, here the number of letters subject i produces. Further, one differentiates *discretionary* from *exogenous* variables. A discretionary variable is one that affects production and that the individual can influence, which in this experiment is the subject's input or letters produced. An exogenous variable is one that the person cannot reasonably influence but that may have an impact on output. The exogenous variable in the experiment was the per-letter money credit, denoted p_i , that was assigned by the experimenter and was unrelated to any working conditions, actions or decisions of the subjects. Thus, i 's output equals the product of i 's per-letter money credit and the number of letters produced by i , that is, $q_i = p_i \cdot x_i$. In certain cells members of a pair were arbitrarily assigned different per-letter credits, which then affected their output, or total money credits.

² Some important contributions to equity theory include George C. Homans (1958), Peter M. Blau (1964), J. Stacy Adams (1965), Elaine Walster et al. (1973), and Reinhard Selten (1978).

The entitlement is posited to relate to discretionary and exogenous variables in accordance with the following principle.

ACCOUNTABILITY PRINCIPLE: *The entitlement varies in direct proportion to the value of the subject's relevant discretionary variables, ignoring other variables, but does not hold a subject accountable for differences in the values of exogenous variables.*

That is, *ceteris paribus*, the entitlement of a subject is proportionate to his/her relevant discretionary variables, relative to others. For example, a subject who produces twice as many letters as his or her counterpart is, all else equal, deserving of twice as much money. Nevertheless, the subject is neither rewarded nor punished for exogenous variables, even if they have an effect on output. For instance, a subject who produces the same number of letters as his or her counterpart deserves the same reward even if the subject's total money credit is twice as much because of an arbitrary difference in the per-letter credits across subjects. In other words, this principle proposes that, for allocation purposes, subjects be held accountable only for factors they can reasonably influence. Of course, different interpretations of what constitutes discretionary or exogenous variables may sometimes be expected. Nevertheless, the results of this and previous studies suggest considerable agreement on the accountability principle and on the proper classification of variables as discretionary or exogenous in a wide variety of circumstances.

One of the treatment variables in the experiment focused on predictions of this principle. This informed the first-phase procedures during which subjects prepared the letters. To test the relevance of discretionary variables and their proportionality to fair allocations, in one treatment each subject was assigned the same 50-cent credit for each letter produced in a five-minute period (i.e., $p_1 = p_2 = 0.50$). All were given materials for 20 letters, more than any could complete in the time given. This was the "discretionary differences" treatment in which fair allocations are predicted to be proportional to inputs or letters produced. To test the irrelevance of exogenous variables, subjects in other sessions were given materials for only ten let-

ters and seven minutes, time enough for all to complete all ten letters. In this treatment, however, members of a pair were arbitrarily assigned different per-letter credits ranging from 25 to 75 cents, but always averaging 50 cents per pair (i.e., $p_1 \neq p_2$ and $(p_1 + p_2)/2 = 0.50$). This was the "exogenous differences" treatment for which fair allocations are predicted to be equal. Thus, in both treatments, the total money credits attributed to members of a pair typically differed, but this variation was solely the result of discretionary differences in the first version and to exogenous differences in the second one.

The entitlement may be expressed algebraically.³ Let \bar{y} represent the total earnings allocated to a pair. Since all money generated by a pair was distributed to that pair, this equals the total output or $\bar{y} = \sum_{i=1}^2 q_i$. Similarly, let \bar{x} denote the total input, or letters produced, by a pair, i.e., $\bar{x} = \sum_{i=1}^2 x_i$. Then the fair allocation may be written

$$(1) \quad \eta_i = \frac{x_i}{\bar{x}} \cdot \bar{y}.$$

Now we turn to the other component of the theory and the structure within which it was tested.

³ The equation here is a simplified version of the original formula introduced in Konow (1996) that consists of the three terms. First, the endowment term is suppressed here. This term represents i 's portion of the allocated variable (here money) that is unrelated to any productive or other activity, which in the experiment was the show-up fee each subject received. Accountability implies that both subjects should receive equal show-up fees. It is dropped from explicit consideration because, in these experiments, all subjects did receive equal show-up fees, were informed of that fact, and, moreover, never made decisions affecting these values. Second, in the more general model, a subject's input may be a function not only of discretionary variables, as here, but also of certain exogenous variables, e.g., the exogenous personal characteristics of the subject, which may affect his or her production. This may require adjusting the input to excise the exogenous factors. In the current experiment no information is provided, or even available, about these exogenous variables as they relate to the subjects' task. Based on the evidence about subject assumptions under such circumstances from the earlier study, no adjustment in inputs is necessary. Third, the fair costs of production that appeared in the original formula vanish here because subjects were not required to bear any money costs for the productive task in this experiment.

B. A Model of Cognitive Dissonance

If fairness were the only concern of people, our task would be considerably facilitated, but, as pointed out earlier, the evidence suggests that people are not so obliging. In particular, their self-interest may conflict with fairness and influence their behavior. Bolton (1991) and Matthew Rabin (1993) have proposed ways of integrating fairness into decision making. This paper suggests another approach based on cognitive dissonance theory [see, for example, Festinger (1957), Philip G. Zimbardo (1969), Elliot Aronson (1976), and Robert A. Wicklund and Jack W. Brehm (1976)]. This social psychology theory is concerned with relations among “cognitions,” i.e., desires, beliefs, opinions, attitudes, or pieces of knowledge. When two cognitions are inconsistent, they are said to be “dissonant,” e.g., the desire to have all the money and the wish to divide it fairly in the dictator experiment. The agent is motivated to reduce dissonance and may, generally speaking, do so by altering behavior, e.g., when the dictator takes less, and/or by changing beliefs, e.g., when the dictator believes it is fair to take more than the fair amount. This section presents a model of cognitive dissonance as applied to three variations, or treatments, of the dictator experiment. As elaborated below, the treatments differ according to whether the dictator allocates between him- or herself and one other person (the standard dictator case), between two other individuals, or does both. The model informs the design of the second phase of the experiment and predicts how the results of the three treatments reveal the entitlement, the degree of self-interest, and the degree of self-deception.

The Standard Dictator Treatment.—We begin with the standard dictator method of allocation, which here involves one subject during the second phase of the experiment deciding unilaterally the division of the money credited in the first phase to his or her joint account. In this treatment, the dictators are in Room A and the recipients in Room B. The theoretical model employs Rabin’s method of incorporating separate terms for cognitive dissonance and self-

deception costs into the utility function.⁴ Suppressing now subscripts for persons, $y \in [0, \bar{y}]$ denotes the amount of earnings that the dictator allocates to one of the subjects, in this treatment, to him- or herself. The dictator’s material utility from this allocation is denoted $v(y)$ where $v(y)$ is assumed twice continuously differentiable with $v_1(y) > 0$ and $v_{11}(y) < 0 \forall y \in [0, \bar{y}]$. Subscript i represents the (partial) derivative of a function with respect to its i th argument with double subscripts signifying second-order (partial) derivatives.

Let $\phi \in [0, \bar{y}]$ represent the amount that the dictator believes it is fair to allocate to the same subject, in this treatment, to him- or herself. If the dictator allocates an amount different from what he/she believes is fair, the dictator may experience cognitive dissonance, e.g., a dictator who takes more than what he/she believes is fair may experience some displeasure at being unfair to his/her counterpart. This is represented by $f(w, \alpha)$, $w \equiv y - \phi$, where $\alpha \in [0, 1]$ is a parameter that indexes the family of functions, $f(\cdot)$, and reflects sensitivity to cognitive dissonance. A higher α represents greater sensitivity and may vary across dictators as well as according to the context, e.g., the procedures of an experiment. It is assumed that $f(\cdot)$ is continuously differentiable in α , $\alpha \neq 1$, twice continuously differentiable in w , $\alpha \neq 1$, and that $f(0, \alpha) = 0 \forall \alpha$. When $\alpha = 0$ cognitive dissonance is completely absent, i.e., $f(w, 0) = 0 \forall w$. At the other extreme, if $\alpha = 1$ then $f(w, 1) \equiv \infty$, $w \neq 0$, that is, it is prohibitively unpleasant to take any amount believed to be unfair. In the intermediate range, $\alpha \in (0, 1)$ whereby $f_1(w, \alpha) \cdot w > 0$, $w \neq 0$, and $f_{11}(w, \alpha) > 0$. That is, f is a strictly convex function of w , meaning that dissonance increases at an increasing rate as the amount taken deviates from what is believed to be fair. In general, for all $w \neq 0$, $f_2(w, \alpha) >$

⁴ Cognitive dissonance theory was formally introduced to economics by George A. Akerlof and William T. Dickson (1982) and was developed more recently by Rabin (1994). The optimization problem here differs from Rabin’s in the following respects: (i) the “moral” amount (the entitlement) is nonnegative and not constrained to zero, (ii) the agent faces a constraint on the amount consumed of the good (here the allocation that the dictator takes) and on the amount that he or she believes to be moral (or fair), and (iii) both the degree of cognitive dissonance and the cost of self-deception are parameterized.

0, $f_{12}(w, \alpha) > 0$, and $\lim_{\alpha \rightarrow 1} f(w, \alpha) = \infty$, that is, absolute and marginal dissonance are increasing in α and approach the $\alpha = 1$ case as α approaches 1. Of course, it follows from these assumptions that $f_1(0, \alpha) = 0 \forall \alpha \in [0, 1)$.

If this were the complete description of the optimization problem, the dictator could choose to believe that it is fair to take all the earnings ($\phi = \bar{y}$) in “good” conscience, that is, without any disutility. It is assumed, however, that there is a cost to choosing beliefs that differ from the entitlement, or one’s detached, intellectually honest view of what is fair. For instance, suppose, based on the outcome of a given letter preparation task, that the dictator, as an observer without any stake in the allocation, would favor an equal split. As one of the impacted subjects, however, the dictator would prefer to take all the earnings and to believe that it is fair to do so. People’s beliefs, however, are not arbitrarily pliable: even when mistaken, their beliefs are typically grounded on and reconciled with some knowledge or experiences. When dictators change beliefs about what is fair, it is assumed to be costly. This may take the form of a costly search for arguments to justify an adjustment in beliefs as well as the displeasure occasioned by such self-serving rationalization. This cost of self-deception is assumed to be a function of the difference between the dictator’s belief and his/her entitlement and is represented $c(z, \beta)$, $z \equiv \phi - \eta$, where $\beta \in [0, 1]$ is a parameter that indicates how costly self-deception is and may vary across dictators and contexts, e.g., with experimental procedures. $c(\cdot)$ is functionally equivalent to $f(\cdot)$.⁵

The standard dictator’s objective function is then assumed to consist of these three terms: the material utility less the cognitive dissonance and deception cost terms. Thus, the dictator chooses the levels of two variables, how much of the earnings to take, and how much to believe it is fair to take, to solve the following problem:

$$(2) \quad \begin{aligned} & \text{Max}_{y, \phi} u(y, \phi, \eta, \alpha, \beta) \\ & \equiv v(y) - f(y - \phi, \alpha) - c(\phi - \eta, \beta) \\ & \text{subject to } y \leq \bar{y}, \phi \leq \bar{y}. \end{aligned}$$

The strict concavity of $v(y)$ and strict convexity of $f(y - \phi, \alpha)$ and $c(\phi - \eta, \beta)$ in y and ϕ ensure the concavity of $u(\cdot)$ in those variables and, therefore, that the second-order conditions for this constrained optimization are satisfied.

The propositions in this section assume the following conditions hold with regard to the values of three parameters: (i) $\alpha > 0$, (ii) $\beta > 0$, and (iii) $\eta < \bar{y}$. That is, dictators experience at least a little disutility from cognitive dissonance and self-deception and the entitlement is less than total earnings. These conditions help simplify the presentation and the proofs (which are available on request from the author). Nevertheless, with greater tedium, all of the propositions that follow can be shown to hold under some set of weaker conditions [usually just condition (i) or (ii)].

Solving and interpreting the first-order Kuhn-Tucker conditions for this problem lead to certain conclusions about the standard dictator’s optimal allocation y^* and the optimal belief ϕ^* , as expressed in the following proposition.

PROPOSITION 1: *For the standard dictator, $\eta \leq \phi^* \leq y^* \leq \bar{y}$. Specifically, the following cases may be distinguished.*

- A. *Complex:* If $\alpha \in (0, 1)$ and $\beta \in (0, 1)$, then $\eta < \phi^* < y^* \leq \bar{y}$.
- B. *Self-deceptive:* If $\alpha = 1$ and $\beta \in (0, 1)$, then $\eta < \phi^* = y^* \leq \bar{y}$.
- C. *Selfish:* If $\alpha \in (0, 1)$ and $\beta = 1$, then $\eta = \phi^* < y^* \leq \bar{y}$.
- D. *Fair:* If $\alpha = \beta = 1$, then $\eta = \phi^* = y^* < \bar{y}$.

In other words, the standard dictator will believe it is fair to take at least his/her entitlement, and possibly more, and will take at least his/her belief, and possibly more, possibly the total earnings. Depending on the dictator’s sensitivity to cognitive dissonance and self-deception, these weak inequalities may be converted to strict equalities or strict inequalities as in the four stated cases. These allow one to infer the

⁵ That is, it is assumed that $c(\cdot)$ is continuously differentiable in β , $\beta \neq 1$, twice continuously differentiable in z , $\beta \neq 1$, and that $c(0, \beta) = 0 \forall \beta$, $c(z, 0) = 0 \forall z$, and $c(z, 1) = \infty, z \neq 0$. Also, for $\beta \in (0, 1)$, $c_1(z, \beta) \cdot z > 0, z \neq 0$, and $c_{11}(z, \beta) > 0$, and for all $z \neq 0$, $c_2(z, \beta) > 0, c_{12}(z, \beta) > 0$, and $\lim_{\beta \rightarrow 1} c(z, \beta) = \infty$.

dictator's type (i.e., α and β values as distinguished in parts A, B, C, or D of the proposition) from his/her choice of ϕ and y values.

The behavior of the so-called *complex* dictator of part A reflects some degree of both selfishness and self-deception. Material utility draws the dictator to take all the earnings, whereas cognitive dissonance pulls in the direction of perceived fairness. The possibility of self-deception allows the dictator to reduce dissonance by choosing a belief ϕ^* greater than η , while creating some self-deception costs. At the optimum, marginal dissonance $f_1(y^* - \phi^*, \alpha)$ equals marginal self-deception costs $c_1(\phi^* - \eta, \beta)$ and, for an interior optimum ($y^* < \bar{y}$), equals marginal material utility $v_1(y^*)$. The *self-deceptive* dictator in part B, for whom dissonance costs are prohibitive, believes it is fair to take more than the entitlement but takes no more than his/her belief. On the other hand, the *selfish* dictator in part C, who has prohibitive self-deception costs, chooses a belief equal to the entitlement, but takes more than his/her belief. The *fair* dictator in part D has both prohibitive dissonance and self-deception costs and, therefore, takes his/her belief, which equals the entitlement.

Proposition 2 states how, in the standard dictator case, the optimal allocation and belief vary with the values of the parameters.

PROPOSITION 2: *For the standard dictator, y^* and ϕ^* vary with η , α , and β as follows.*

- A. $\partial y^*/\partial \eta \geq 0$ and $\partial \phi^*/\partial \eta \geq 0$, with strict inequalities for $y^* < \bar{y}$ and $\phi^* < \bar{y}$, respectively.
- B. $\partial y^*/\partial \alpha \leq 0$ and $\partial \phi^*/\partial \alpha \geq 0$, with strict inequalities for $\eta < y^* < \bar{y}$ and $\eta < \phi^* < \bar{y}$, respectively.
- C. $\partial y^*/\partial \beta \leq 0$ and $\partial \phi^*/\partial \beta \leq 0$, with strict inequalities for $y^* < \bar{y}$ and $\phi^* < \bar{y}$, respectively.

Proposition 2A means that, except for corner solutions, a higher entitlement reduces the cost of self-deception and leads to the belief in a higher fair amount, which then reduces dissonance and causes the dictator to take more. Proposition 2B states that greater sensitivity to dissonance lowers selfishness but increases self-deception, except for corner solutions. That is, being more sensitive to fairness encourages the dictator to be less "selfish"

but to convince him- or herself that a higher allocation is justified. According to Proposition 2C, higher self-deception costs reduce self-deception and therefore selfishness except, as before, for corner solutions.⁶

The combined model of fairness and cognitive dissonance generates predictions about the values of the entitlement, the chosen belief about fairness, and the actual allocation. Of these three values, however, the standard dictator experiment only reveals the last. The theory, however, suggests two experimental variations, introduced below, to identify the two remaining values.

The Benevolent Dictator Treatment.—If the dictator's stake in the outcome is removed, so also is the inducement to take more than the fair allocation and to deceive, and the dictator's decisions should reflect the entitlement. This study, therefore, introduces a second dictator treatment involving three sets of subjects who do not at any point participate in the standard dictator experiment. Two groups of subjects, in Rooms labeled A and B, perform the first-phase task as in the standard version. In the second phase the third group, in Room C, acts as *benevolent dictator*: each person in Room C decides for his or her anonymous counterparts in Rooms A and B the allocation of rewards jointly earned by them. The benevolent dictator is paid a fixed amount for this decision unrelated to this decision or the earnings of the counterparts. This dictator is benevolent in the sense that the only concern is for fairness to one's counterparts and honesty to oneself.

Eliminating material utility from the dictator's utility function results in the following maximization problem:

$$(3) \quad \text{Max}_{y, \phi} u_b(y, \phi, \eta, \alpha, \beta) \\ \equiv -f(y - \phi, \alpha) - c(\phi - \eta, \beta) \\ \text{subject to } y \leq \bar{y}, \phi \leq \bar{y}.$$

⁶ The conclusions of Propositions 2B and 2C parallel those of Rabin's (1994) Propositions 1C and 1B, respectively. They differ with respect to method of proof, and here the analysis is extended to include finite constraints on y and ϕ and to the cases of prohibitive dissonance and prohibitive self-deception costs.

Note that in this treatment the dictator is allocating not to him- or herself but to counterparts, and now y , ϕ , and η represent the values chosen or perceived by the dictator with respect to one of the two counterparts (say, Room A; the values for the Room B subject are simply \bar{y} minus each of the respective Room A values). This leads to Proposition 3.

PROPOSITION 3: *For the benevolent dictator, $\eta = \phi^* = y^* < \bar{y}$.*

Quite simply, the benevolent dictator maximizes utility by minimizing the disutility from dissonance and self-deception, that is, by believing the entitlement is fair and by allocating the belief. Since the allocations are real and not hypothetical, and are not obscured by narrow self-interest as in standard dictator games, this treatment provides the most substantive test to date of the fairness theory employed here.

The Double Dictator Treatment.—The remaining variable we wish to quantify experimentally is the standard dictator's belief about what is fair, ϕ^* , which sheds light on the degree of self-deception. A third dictator treatment is introduced in this study that attempts to identify this value by the following means. First, the standard dictator experiment with "exogenous differences" in earnings is conducted as previously described: all subjects prepare the same number of letters and the only difference in money credited to each is the result of arbitrary differences in the per-letter credits. Specifically, in this version, the dictator (in Room A) is always assigned a per-letter credit that is higher than that of the recipient (in Room B). This provides the dictator with a contextual pretense for taking a more-than-fair amount. Immediately thereafter and without prior knowledge of this fact, the standard dictator is put in the position of a benevolent dictator who is facing a new anonymous pair: one subject (in Room C) has the same money credit as the dictator, whereas the other (in Room D) has the same credit as the dictator's earlier counterpart in Room B.⁷ This *double dictator* now chooses

allocations for his Room C *Doppelgänger* and for the Room D subject. If the double dictator's belief is unchanged between rounds, the optimal allocation to the Room C counterpart in the second round, denoted y^{**} , will equal ϕ^* , the dictator's belief about the fairness of his/her own allocation in the first round.⁸ This is because, as a benevolent dictator in the second round, material utility plays no role and, if the belief is stable, self-deception costs are fixed. Thus, the dictator is left only to minimize cognitive dissonance, which is accomplished by allocating the belief.

The double dictator chooses a y and a ϕ , which now denote his/her allocation and belief about what is fair, respectively, to the Room C counterpart, with the stipulation that ϕ is assumed to be the same as the dictator's belief about his/her own fair allocation in the first round, ϕ^* , i.e., that $\phi = \phi^*$. The double dictator's maximization problem is as follows.

$$(4) \quad \begin{aligned} \text{Max}_{y, \phi} \quad & u_d(y, \phi, \eta, \alpha, \beta) \\ & \equiv -f(y - \phi, \alpha) - c(\phi - \eta, \beta) \\ \text{subject to } & y \leq \bar{y}, \phi = \phi^*. \end{aligned}$$

The intuition just stated is formalized in Proposition 4.

PROPOSITION 4: *Assuming stable beliefs, the double dictator allocates his belief, $y^{**} = \phi^*$.*

neutral and contextual. It is designed to create a bias toward the kind of behavior (selfishness) that one seeks here to explain as well as a parallel between the position of the Room A dictator and the Room C counterpart. This is consistent with the studies of Roth and J. Keith Murnighan (1982) and Matthew Spiegel et al. (1994), which suggest that providing information on asymmetries between subjects has an impact on behavior.

⁸The rationale for assuming stable beliefs under the procedural circumstances of this experiment is elaborated in the following section. Note that, theoretically, the double dictator version could also be conducted with discretionary differences in earnings but that two practical problems would arise: it is highly unlikely that very many subjects in Rooms C and D would both prepare exactly the same number of letters as subjects in Rooms A and B as needed, and the dictators' pretense for taking more (higher money credits) would be missing for any Room A subjects who produce less than Room B.

⁷The asymmetric credit favoring the dictator is the only element of the experiments that is intentionally both non-

Differences in Earnings	Recipients	Dictator	
		Room A	Room C
Discretionary	A, B	Standard Dictator - Discretionary	Benevolent Dictator - Discretionary
Exogenous	A, B	Standard Dictator - Exogenous	Benevolent Dictator - Exogenous
	C, D	Double Dictator - Exogenous	

FIGURE 1. EXPERIMENTAL DESIGN

The experimental design described in this section, and summarized in Figure 1, involves two treatment variables. The one, the *differences in earnings* variable, involves two variations, the discretionary differences and the exogenous differences treatments. The other, the *dictator* variable, comprises two elemental versions, the standard (Room A) dictator and benevolent (Room C) dictator, plus the double dictator. In this experiment, all of the standard dictators in the exogenous differences treatment participated in a double dictator round. The basic design is, therefore, a 2×2 appended by the double dictator cell in which Room A standard dictators allocate as benevolent dictators to Rooms C and D. The following section explains the particulars of the experiment.

II. Experimental Procedures

The 360 participants in this experiment were recruited from among diverse undergraduate classes at Loyola Marymount University (or LMU) and from the LMU Economics and Psychology Subject Pools. No one was informed of the content or purpose of the experiment, only that the participants “may be asked to perform a simple task, ... to make a decision” and “may receive some money” in the general range of actual amounts. Subjects were permitted to participate only once.⁹ With the average session

⁹ Also, all rooms in a session were conducted simultaneously, except for the double dictator treatment. Those sessions, which required four rooms, were conducted at two times, first Rooms A and B and later Rooms C and D, in consideration of logistics (i.e., time and room availability).

lasting about 50 minutes and paying approximately \$8.50, mean hourly compensation was about \$10 per hour, which comfortably exceeds the opportunity cost of most LMU students. Moreover, after participating and being paid, 88 percent of the subjects responding indicated that they would like to be called again to participate in experiments.

Dictators and recipients showed up to separate, preassigned meeting rooms and received the same \$3 show-up fee in cash. Dictators always numbered 12 per session, and multiple sessions generated 36 observations each in the standard/exogenous and double/exogenous cells and 24 observations in each of the other three cells. All participants were seated at portable study carrels that shielded their decisions and responses from the view of others.

Next, subjects were given copies of the phase 1 instructions that explained the anonymity conditions and the phase 1 task: subjects fold letters, stuff them into envelopes, and place them through a slot in a sealed box.¹⁰ The letters were in a typical format used by the University Relations office at LMU to attract funds. No one was told that these letters would or would not be used for any purpose and no subject asked, but no one expressed any doubt that the letters were authentic (in fact, the private comments to me of one of the assistants indicated that even he regarded them as genuine). Thus, participants could be expected to take their contribution as serious productive activity. Otherwise, subjects were given complete information about

¹⁰ The complete instructions are available on request from the author.

the structure (but not the purpose) of the experiment.¹¹

After the task was complete, letters were counted and this information was taken to the experimenter in the dictator room. For the exogenous differences sessions, subjects were assigned counterparts based on a previously determined random matching. For the discretionary differences sessions, subjects were matched during the experiment using a laptop computer: the high scorer in one room was matched with the low scorer in the other, the second-to-highest scorer with the second-to-lowest scorer, etc. Individual and joint money credits were calculated and recorded on forms to be presented to the subjects.

Subjects then received the phase 2 instructions that informed all parties that the dictators had “been arbitrarily chosen to decide how the total will be distributed” and told the dictators: “This decision is completely up to you and is confidential: only the experimenter will know who made this decision.” In the exogenous treatments, they also stated that any difference in credits between the subjects “is completely arbitrary: in other words, any difference in this credit does not reflect any difference in the quantity or quality of work by you or your counterpart or any difference in your working conditions.” The Room A (and Room C in the double dictator cell) per-letter credits ranged from 55 to 75 cents, in 5-cent increments, and the corresponding Room B (and Room D) credits ranged therefore from 45 to 25 cents. For a given pair, the average per-letter credit was always 50 cents. In the case of the benevolent dictators (and in their case alone), there was only one phase. They were informed of their role as allocator, of their \$5 fixed fee in addition

to the \$3 show-up fee and, in the exogenous version, of the possibility of an arbitrary difference in credits among their counterparts. Standard and benevolent dictators were then given a form with the results of the task and a space for indicating how the total should be distributed. They had five minutes to make their decision and to put the form in the envelope provided. The envelopes were then collected. While their payments were being prepared confidentially, the dictators were given five minutes to fill out a questionnaire that posed an open-ended question about why they chose the allocation that they did. Except for the double dictators, the subjects then received their payments individually and confidentially, signed a receipt, and left.

In the double dictator version, immediately after returning the questionnaire, Room A dictators received “Further Phase 2 Instructions” regarding the benevolent dictator round. They reiterated the fundamentally equivalent conditions between Rooms C and D and stated that Room A subjects were now to decide the allocations between those rooms for which they would be paid an additional \$3 (an amount judged to be in rough proportion to the additional time expended on this exercise). Room A subjects were given forms with the Room C and Room D results and space for their decision, which were collected after five minutes. Then they responded to a second open-ended questionnaire asking why they allocated as they did in this round while their payments for all rounds were prepared. After five minutes these were collected, and they individually received their payments, signed their receipts, and left.

As stated in the previous section, this double dictator decision is predicted to reveal the standard dictator’s chosen belief of what is fair as long as this belief is stable between rounds. This assumption of stable dissonance-reducing self-deception is supported by social psychology studies, which suggest that successful dissonance reduction is both irreversible (Mark R. Lepper et al., 1970) and long lasting (Danny Axsom and Joel Cooper, 1981). The latter find, in a study of weight loss, that the effects of dissonance reduction not only persist, but may actually increase, after six months and one year. In the current experiment, subjects were confronted with the double dictator decision

¹¹ In two cases, however, certain information was not provided until a later phase. First, Room A and B subjects were told in phase 1 that the money earned by a pair would be distributed in cash to that pair but that the details on this would be provided later. The main goal in this was to avoid differential effort in standard dictator games resulting from different expectations of reward based on who controlled. Second, as already mentioned, double dictators were not told in the standard dictator round that they would be participating in a benevolent dictator round. This was to avoid any strategic choice of beliefs and allocations in the first round in expectation of having to reconcile those choices as a benevolent dictator later.

TABLE 1—SUMMARY OF THEORETICAL AND EMPIRICAL RESULTS

Treatment	Actual values				
	(1) Theoretical allocation	(2) Mean entitlement	(3) Mean allocation	(4) $H_0: (2) = (3)$ t -value	(5) Percentage perfect fit
Benevolent/discretionary	x_a/\bar{x}	0.491	0.494	0.072	45.8
Standard/discretionary	$[x_a/\bar{x}, 1]$	0.516	0.644	2.152**	83.3
Benevolent/exogenous	0.5	0.5	0.508	0.723	87.5
Standard/exogenous	$[0.5, 1]$	0.5	0.592	2.808***	94.4
Double/exogenous	$[0.5, y_a/\bar{y}]$	0.5	0.558	4.205***	86.1

Notes: Allocations and entitlements are expressed as Room A's fraction of the total. Theoretical allocations are the points or, if expressed in brackets, intervals predicted by the theory. The t -value is for a two-tail test of the hypothesis that the mean actual allocation equals the mean actual entitlement.

*/**/***: Reject the null hypothesis at the 10/5/1-percent level of significance.

immediately after justifying their standard dictator decisions on a questionnaire. An important reinforcing factor in the Axsom and Cooper (1981) study was that the justification for dissonance reduction not be forced but freely chosen by the subjects. This was guaranteed in the current study by the completely voluntary decision structure and by a statement read to subjects just prior to the instructions that their participation was voluntary and that they could withdraw at any time. To the extent that the double dictator decision is an imperfect measure, it is probably on the side of underestimating rather than overestimating the degree of self-deception (an argument elaborated in the following section).

III. Results

Since a pair's total input \bar{x} and total earnings \bar{y} vary somewhat in the discretionary treatments, throughout this section individual subject values are stated as fractions to facilitate comparisons across subjects and treatments. Thus, letting x_a represent the letters produced by a Room A subject and y_a his or her dollar allocation, the subject's fractional entitlement and fractional allocation are x_a/\bar{x} and y_a/\bar{y} , respectively. The allocations predicted by the theory advanced in this paper are presented in column (1) of Table 1. With discretionary differences in credits, the optimal allocation in the benevolent dictator cell is the fractional entitlement or x_a/\bar{x} , whereas in the standard dictator version it is in the interval $[x_a/\bar{x}, 1]$. When

differences in credits are exogenous, the benevolent dictator is expected to allocate 0.5 of the total to each, whereas the standard dictator should take a fraction on the interval $[0.5, 1]$. The optimizing double dictator allocates somewhere between half and what he took in the earlier standard dictator round or an amount on $[0.5, y_a/\bar{y}]$.

Mean fractional entitlements (or inputs) to Room A appear in column (2) of Table 1. These, incidentally, are not significantly different from 0.5 for the discretionary treatments and, of course, equal 0.5 by design for the exogenous treatments.¹² The mean fractional allocations, which appear in column (3), provide favorable evidence on the theory. Note that these are not significantly different from mean fractional entitlements for the two benevolent dictator treatments as suggested by the two-tail t -values in column (4). The mean allocations in the three other cases, however, are significantly greater than mean entitlements, revealing that standard dictators are taking, on average, more than the entitlement and are engaging in self-deception. Also, in comparison to their benevolent coun-

¹² Nevertheless, in all treatments earnings spanned a wide range, and average differences were not insubstantial. In the discretionary treatments, the average high earner (whether in Room A or B) produced and earned 52 percent more than the average low earner. In the exogenous treatments productivity was equal but, because of different per-letter credits, the earnings of the average high earner were 87 percent greater than those of the average low earner.

terparts, standard/discretionary dictators take more ($t = 2.743$; $p = 0.004$), as do standard/exogenous and double dictators ($t = 2.024$; $p = 0.024$, and $t = 2.571$; $p = 0.006$, respectively). The double dictators allocate on average less to Room C than they do to themselves as standard dictators, but for the total sample this difference is not significant ($t = 0.940$; $p = 0.175$). This is the result of the high coincidence of first- and second-round allocations, partly because of self-deception but mostly because of numerous fair allocations in both rounds.¹³ Looking only at the unfair subset, the mean allocation for the standard/exogenous treatment is 0.732 and for the double/exogenous 0.591. These allocations are both significantly different from 0.5 ($t = 5.683$; $p < 0.001$, and $t = 4.510$; $p < 0.001$, respectively) and from one another ($t = 3.095$; $p = 0.002$). The implication is that, on average, unfair dictators are like complex dictators: they take more than what they believe is fair and believe it is fair to take more than the fair amount. Finally, column (5) of Table 1 states the percentage of allocations that conform exactly to the theory, i.e., the percentage of observations that equal the point predictions in the case of the discretionary treatments or the interval predictions in the case of the standard and double treatments.

Proceeding now to a visual review of the individual results, fractional allocations are measured on the vertical axes of Figures 2A to 2D. For the discretionary treatments depicted in Figures 2A and 2B, the dictator's fractional input x_a/\bar{x} appears on the horizontal axis. For the exogenous treatments of 2C and 2D, the dictator's credit p_a appears on the abscissa. The fractions allocated to Room A subjects are indicated with asterisks (*). The theory predicts that benevolent dictators with discretionary dif-

ferences in credits will allocate fairly, that is, in proportion to inputs as represented by the 45-degree line in Figure 2A. Standard dictators with discretionary differences in credits are expected to take an amount equal to or greater than the fair amount, i.e., an allocation on or in the shaded area above the 45-degree line in Figure 2B. When differences in credits are exogenous but inputs equal, benevolent dictators should allocate equally, that is, on the horizontal line at 0.5 in Figure 2C. Under the same circumstances, however, standard dictators will take amounts on or in the shaded area above the horizontal line in Figure 2D. Finally, when these same standard dictators become benevolent dictators, their double dictator allocations are expected to equal the fair allocation of 0.5, their earlier allocations to themselves, or some amount between these two values. Where standard and double dictator allocations are equal (which is the case for 63.9 percent of these observations), a single asterisk represents both allocations of a subject. Otherwise, the asterisk represents the standard dictator allocation and a dotted line extends to that subject's double dictator allocation.

A visual inspection of Figure 2 suggests support for the theory. The benevolent dictator treatments yield results close to the point predictions of the theory. In the standard dictator cases, on the other hand, the predictions of the theory are, in a sense, weaker since it predicts a range of possible outcomes. In these there is an equal number of fair dictators and ones who take more than the fair amount (with another 10 percent taking less than the entitlement). If either self-interest or fairness routinely dominated, one might argue for a more parsimonious model with more specific predictions for these treatments. Nevertheless, the implication of the results from this and numerous prior standard dictator experiments is that the phenomenon to be explained is precisely the dispersion in outcomes within a specific range.

The tests of differences in means reported in Table 1 are consistent with the theory. Nevertheless, Figures 2A and 2B highlight certain more-specific predictions of the theory for the discretionary treatments. In particular, allocations are predicted to lie either on or on and above the 45-degree line where allocations equal entitlements. We turn, therefore, to a

¹³ Interestingly, fully 45 percent of standard dictators, including discretionary and exogenous versions, allocate the fair (but not necessarily equal split) amount, which is higher than the percentage of equal (and presumably fair) splits found in most other dictator experiments [the Kahneman et al. (1986a) study with 76-percent equal splits is an exception]. It may be that, in the current experiments, accountability issues are more tangible and unfairness, therefore, more unpleasant.

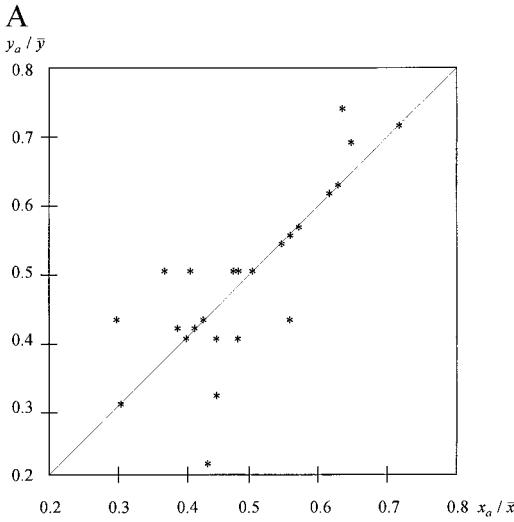


FIGURE 2A. BENEVOLENT/DISCRETIONARY CELL

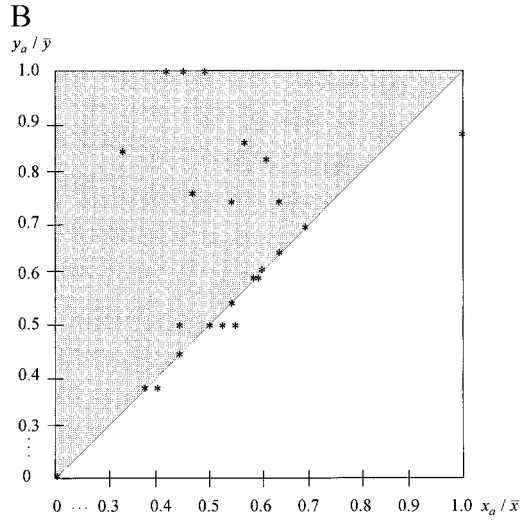


FIGURE 2B. STANDARD/DISCRETIONARY CELL

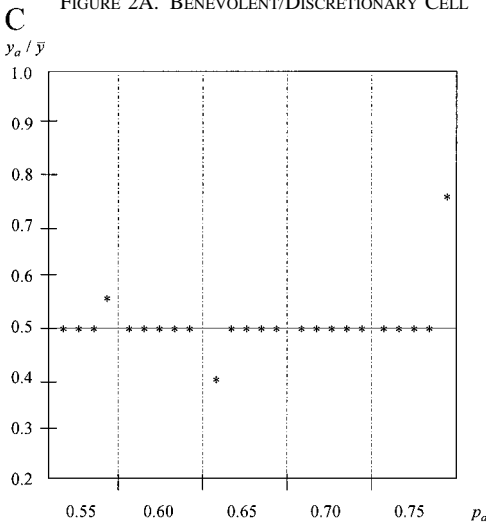


FIGURE 2C. BENEVOLENT/EXOGENOUS CELL

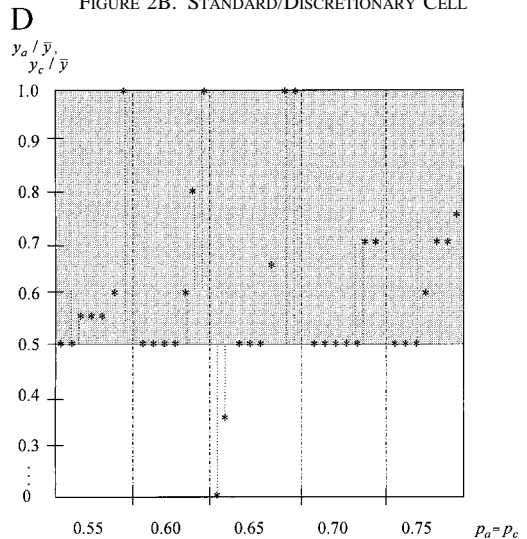


FIGURE 2D. STANDARD/DOUBLE/EXOGENOUS CELL

paired difference test, i.e., we examine the significance of deviations of individual dictator allocations from the entitlement, or $y_a/\bar{y} - x_a/\bar{x}$.¹⁴ Based on the theory, one would expect these deviations to be significantly different from zero for standard dictators but not for benevolent dictators. For the exogenous treatments, the entitlement is constant (0.5), and the paired difference test is equivalent to the test of

differences in means already reported in Table 1, but for the discretionary cases this is a distinct test. The first column of Table 2 summarizes these mean differences and the corresponding *t*-values for the discretionary treatments. As indicated there, the difference is not significant for the benevolent/discretionary case but is significant for the standard/discretionary case, consistent with the theory and with the results on differences in means.

Table 2 presents additional support for the theory from ordinary least-squares (OLS) re-

¹⁴ This test was suggested by a referee.

TABLE 2—ANALYSIS OF DISCRETIONARY TREATMENTS

Treatment	Paired difference test	OLS regressions				R^2
	Mean difference $y_a/\bar{y} - x_a/\bar{x}$ (<i>t</i> -statistic)	Parameter estimate (standard error)		Null hypothesis <i>t</i> -statistic		
		$\hat{\gamma}_0$	$\hat{\gamma}_1$	$\gamma_0 = 0$	$\gamma_1 = 0$	
Benevolent/discretionary	0.003 (−0.156)	0.045 (0.076)	0.914 (0.150)	0.593	6.078***	0.627
Standard/discretionary	0.128 (2.959)***	0.293 (0.138)	0.679 (0.254)	2.126**	2.672**	0.245

*/**/***: Significant at the 10/5/1-percent level.

gressions. The regression equation for the *i*th Room A subject in the discretionary treatments is

$$(5) \quad y_i/\bar{y} = \gamma_0 + \gamma_1 x_i/\bar{x} + \varepsilon_i,$$

where ε_i is an error term. For the benevolent/discretionary treatment, Proposition 3 predicts that the intercept of this equation equals 0 and that the slope equals 1. OLS estimates of these parameters (denoted $\hat{\gamma}_0$ and $\hat{\gamma}_1$, respectively) are consistent with the hypothesis: the intercept of 0.045 is not significantly differently from 0 but the slope of 0.914 is. The R^2 equals 0.627, i.e., 62.7 percent of the variance in allocations is accounted for by differences in inputs. Moreover, using an *F*-test, the slope turns out not to be significantly different from 1 ($F = 2.408$, $p = 0.328$).

Turning to the standard/discretionary treatment, an implication of Proposition 2A is that dictators with greater entitlements tend to take larger allocations, that is, that equation (5) has a positive slope. Indeed, the slope estimate of 0.679 in Table 2 is significant at the 5 percent level and suggests that a \$1 increase in the entitlement leads dictators to take, on average, 68 cents more. Moreover, differences in inputs across subjects account for about one-quarter of the variance in allocations.

Now we focus on the standard/double treatment. In this treatment, 86.1 percent of dictators made decisions in both the first and second rounds that are consistent with Proposition 1. Of this group (call it the “conforming subset”), 10 percent are complex (both selfish and self-deceptive) dictators, 29 percent self-deceptive,

16 percent selfish, and 45 percent fair. Thus, 39 percent of the conforming subset self-deceive to some extent. More important, among unfair dictators in this group, i.e., among those predicted by the theory to be susceptible to self-deception, 71 percent do self-deceive. For unfair dictators in the conforming subset, the percentage of unfairness resulting from self-deception, measured individually as $[(y_c - \eta_a)/(y_a - \eta_a)] \cdot 100$ percent and averaged across the group, is 57 percent. Interestingly, comments of dictators provided later on questionnaires tend to substantiate this interpretation. Despite the fact that most of the more detached benevolent dictators view the credit differences as irrelevant, many of the standard dictators choose to believe otherwise, including a self-deceptive dictator who writes: “For some reason I believed that I was getting more money for a reason.”

Several arguments may be made why y_c provides a conservative estimate of the degree of self-deception. First, to whatever extent the assumption of a stable ϕ is inaccurate, y_c tends to be less than ϕ in the first round. This is because the double dictator is motivated to reduce self-deception costs by readjusting his ϕ toward the entitlement in the second round. Second, even if ϕ is stable between rounds and y_c accurately gauges self-deception in the first round, it may understate the degree of future self-deception. This is the result of the effect, identified by Axsom and Cooper (1981) and mentioned in the previous section, for the effects of dissonance-reducing measures to increase rather than decrease over time. In the current context, complex and self-deceptive dictators might, for example, think of more items to add to their list of rationalizations.

Third, some self-deception might take a form that is not captured by the proposed measure. This is probably the most significant manner in which self-deception is underestimated here. Almost one-quarter of the unfair dictators expressed doubt in the questionnaires about whether their Room B counterparts really existed. As one puts it: "I do not believe the other room or my counterpart existed. Thus, any portion not given to me would be wasted." Either these dictators are sincere in their doubt, or they are engaging in a kind of self-deception on this point. In either case, y_c tends not to capture otherwise genuine self-deception since selfishness in the first round and fairness in the second round come cheap. As the subject just cited continues after the second round: "At this point whether I believe in the existence of rooms C or D is meaningless. I do not stand to profit in excess of my 3 dollars; therefore I am free to provide an equitable sum for both." If the admitted skeptics are eliminated from the sample, average self-deception as a percentage of unfairness rises from 57 to 71 percent.¹⁵

In general, the reasons given by different types of dictators coincide with what would be expected of their type. For example, one complex dictator both acknowledges some unfairness while confirming the thoroughness of the self-deception by writing after the second round: "I can judge fairly in situations where I am not affected by the decision." A selfish dictator, whose type is unconditionally self-interested, writes after the first round: "I took

advantage of the position I was put in." After the second round the same subject writes: "I have no reason to slight [Rooms] C and D. I divided it equally between them because since I have no personal contact with either of them, I have nothing to benefit by giving extra to one and slighting the other." Justice was explicitly on the minds of many of the standard/double dictators. Fairness, equity, equality (which is equivalent in this treatment to fairness), or one of their cognates is mentioned by 63 percent of fair dictators and by 18 percent of unfair dictators.

Consider now the effect of differences in per-letter credits on allocations in the exogenous treatments, given by the following regression equation:

$$(6) \quad y_i/\bar{y} = \gamma_0 + \gamma_1 p_i + \varepsilon_i.$$

A higher per-letter credit might be viewed by unfair dictators as greater justification for taking more than the fair amount. If so, a higher per-letter credit reduces the cost of self-deception, i.e., it reduces β . Then, according to Proposition 2C, dictators with larger credits would allocate more to themselves and to their Room C counterparts. That is, the slope of equation (6) would be positive for both Room A and Room C allocations. The results of OLS regressions indicate that the slope for standard/exogenous dictators is insignificant and the wrong sign ($\hat{\gamma}_1 = -0.143$, standard error = 0.474), whereas the slope for double/exogenous dictators is positive and significant at the 5-percent level ($\hat{\gamma}_1 = 0.429$, standard error = 0.188). The latter estimate suggests that a \$1 increase in the money credited to the Room A dictator results in a 43-cent increase in self-deception. Thus, if the assumed relationship between credits and β applies, the evidence on Proposition 2C is partly favorable and partly inconclusive.¹⁶

¹⁵ An additional argument is that presenting decision makers with a problem identical to a previous one with a single variation may invite a different response. Evidence in the benevolent/discretionary condition of such a desire by subjects to avoid the most obvious choice is discussed below. To the extent that first-round allocations to Room A are high, second-round allocations to Room C might, therefore, tend to be lower. The opposite point was suggested to the author by Werner Güth: decision makers might, out of a desire for consistency, tend toward the same choice in the second round. Along similar lines, a referee suggested that, given experimenter knowledge of subject identity, subjects might match their earlier allocations to maintain the appearance of consistency to the experimenter. The dictator decisions themselves do not provide clear evidence either way: unfair dictators were about evenly split between those who chose the same allocation to their Room C counterparts and those who chose less.

¹⁶ One problem with the sample on which these estimates are based is that the subjects who do not believe their counterparts exist are, as explained earlier, inclined to take all of the earnings but to allocate equally to Rooms C and D and, therefore, not expected to respond to differences in credits. Eliminating them from the sample results in a positive slope of 0.204 for standard dictators (standard error = 0.387), although it is not significant. Using this subgroup,

Examining Table 1 and Figures 2A–2D, it is noteworthy not only that the results are generally consistent with the theory, but also that a large fraction of the data fit perfectly the allocations predicted by Propositions 1, 3, and 4. This is less surprising for the treatments in which the theory predicts ranges, but even in the benevolent dictator treatments many observations match the point predictions. Nevertheless, this percentage is noticeably lower for the benevolent/discretionary case than for the benevolent/exogenous case. Here many subjects seem to be allocating in proportion to inputs, plus or minus a little. Additional sessions were conducted that suggest that this dispersion is an experimental artifact. In the benevolent/discretionary—version 2 case new Room C dictators were presented with the same Room A and Room B results as in the original version, but the decision forms were slightly different. Originally, the form provided results for the number of envelopes completed and the total money credited *separately* for each counterpart. The version 2 form provided separate information only on the number of envelopes completed by each counterpart; the money credit appeared only as a total for the pair. Thus, to allocate fairly, this second version required dictators to calculate the individual dollar amounts on their own, instead of merely transferring the values already provided. The suspicion behind this changed format was that, although benevolent dictators are primarily concerned with fairness, they are also averse to doing the obvious. This aversion is a problem for the original benevolent/discretionary but not for the benevolent/exogenous and benevolent/discretionary—version 2 treatments. The results of version 2 are presented graphically in Figure 3. Comparing Figures 2A and 3, the reduction in dispersion is striking. The number of decisions that exactly fit the point predictions of the theory jumps from 45.8 to 79.2 percent. Regressing fractional allocations on fractional inputs as before using equation (5), the intercept of -0.015 is not significantly different from 0 nor is the slope of 1.046 significantly different from 1. The R^2

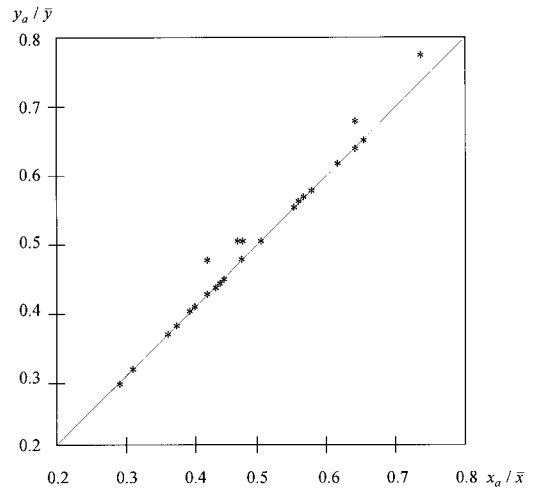


FIGURE 3. BENEVOLENT/DISCRETIONARY—VERSION 2

rises from 0.627 in the original benevolent/discretionary sessions to 0.983 in version 2.

IV. Discussion

This section provides an overview of the literature related to this study beginning with a discussion of research into biases in fairness judgments. Then I review some other dictator experiments and argue that the theory proposed here performs well versus competing theories in explaining, organizing, and reconciling these results.

There is an extensive literature among social psychologists that is related to many of the concepts and tests proposed here and raised in the recent economics literature on bargaining. One prominent and early contributor to this literature is Gerold Mikula (1972a, b; also Mikula and Hans Uray, 1973), who has also conducted experiments aimed at determining how individuals allocate jointly earned rewards. His results are generally consistent with the fairness theory used in this paper, although he finds a *generosity* bias by dictators, which contrasts with the *egocentric* bias seen here and otherwise identified by researchers (e.g., Messick and Sentis, 1979; Spiegel et al., 1994). This is probably an artifact, as he himself implies, of experimental procedures including his use of deception, fictional counterparts, nonproductive tasks, and unspecified and small rewards.

the double dictator relationship becomes somewhat stronger and more significant.

Among the most thorough research into egocentric, or self-serving, biases in fairness judgments is that of Linda Babcock, George Loewenstein, and their collaborators. In a series of experimental investigations, they have examined the behavior of subjects faced with a contextually rich tort case based on an actual trial. In the first of these studies (Loewenstein et al., 1993), subjects were initially assigned randomly to the role of plaintiff or defendant, then were provided with identical details about the case, and finally attempted to negotiate a settlement that determined their actual payments. Delays in settlement caused the parties to incur penalties, and failure to settle voluntarily in the time provided resulted in a settlement based on the actual judge's decision (which was previously unknown to both parties). After reading the materials but before negotiating, subjects were asked to indicate their predictions of the judge's award as well as of what they considered fair. Following negotiation, subjects were asked to recall and rate the importance of arguments favoring both the plaintiff and the defendant. Plaintiffs and defendants were found to have self-serving estimates of the judge's award and of what was fair. In addition, they recalled more arguments favoring their own position and weighted those arguments more heavily. Significantly, parties that settled out of court expressed more similar assessments of the judge's award and of what was fair, and exhibited a lower incidence of egocentric recall of arguments and of egocentric rating of their importance.

Babcock et al. (1995) set out to test whether the self-serving bias *causes* bargaining impasse as opposed, say, to both phenomena being caused by a third factor such as a personality trait. To this end, they followed the basic design of their earlier study, introducing a manipulation of the self-serving bias. Specifically, in the control condition subjects were informed of their role as plaintiff or defendant prior to reading the materials and to offering their estimates of the judge's decision and of the fair settlement, as in the previous experiment. In the experimental condition subjects were informed of their roles only after reading the materials and estimating the judge's and fair settlements (but, of course, prior to negotiation). As predicted, subjects in the experimental condition

were significantly more likely to settle and to do so in less time. In addition, they were significantly more likely to agree on the judge's and fair settlements and less likely to differ in the importance they attached to self-serving arguments. Of course, the timing of information about parties' roles may generally not be manipulated outside the laboratory, but Babcock and Loewenstein (1997) report on other possible interventions to reduce the self-serving bias and resultant impasse. They find that informing subjects who already know their roles of the existence of the self-serving bias and having them list the weaknesses in their own case significantly decreases differences in their estimates of the judge's award and in the occurrence of impasse.

In response to questions about the applicability of their experimental findings to bargaining in a real-world setting, Babcock et al. (1996) examine actual teacher contract negotiations. Labor union and school board negotiators are found to exhibit a self-serving bias in their selection of comparison groups. Moreover, strike activity is positively related to the magnitude of difference in salaries between the union and board lists of comparable school districts. Babcock, Loewenstein, and their collaborators mount compelling evidence from the laboratory and the field of egocentric biases in fairness judgments. Their observations are consistent with the self-serving self-deception posited in this paper and with the self-serving processing of information about which Rabin (1995) formulates a theory.

In substance, the current study shares an interest in self-serving biases found in the studies just discussed. In form, however, it is closer to more conventional and less contextually complex dictator experiments to which I now turn. Kahneman et al. (1986a) confronted subjects, a *random fraction* of whom were paid, with two choices: an equal split or one strongly favoring the allocator. From the predominance of equal splits they draw conclusions about the importance of fairness. Forsythe et al. (1994) and Martin Sefton (1992) find that average dictator giving is nontrivial but that paid dictators are considerably less generous to their counterparts than unpaid or randomly paid dictators. With respect to the other feature of the Kahneman et al. (1986a) study, Bolton et al. (1996) find no

significant difference in dictator generosity between a two-choice treatment and one with a greater array of choices. They also examine certain other hypotheses and reject several that differ from the detached and purely distributional concern suggested in the current paper. The one hypothesis that they accept is based on the notion that dictators would rather err in their own favor than in their counterpart's favor, consistent with the theoretical model presented here.

Hoffman et al. (1994, 1996) argue that the wide variation in the degree of generosity by dictators is the result of expectations of reciprocity and a self-interested concern for this social quid pro quo. Specifically, Hoffman et al. (1996) attribute this variation to the effect of different instructional language and procedures on "social distance," i.e., on the perceived proximity to or isolation from social interaction. They conduct six dictator experiments with different language and procedures, which, they claim, capture this effect.

Hoffman et al. (1994, 1996) cleverly trace the source of variation in generosity to differences in experimental procedures and instructional language. Their results are convincing, but the behavioral phenomenon that underlies them is open to interpretation. They attach importance to the "anonymity hypothesis" that attributes giving to the perceived likelihood that the dictator's decision will be known to others. Nevertheless, the results of Timothy N. Cason and Vai-Lam Mui (1997) (and, arguably, of Hoffman et al., 1996) suggest that this effect is not very important, and Bolton et al. (1998) find not only that it is insignificant, but that the shift of offers in their study actually works against the hypothesis.

I believe that the diminished generosity identified by Hoffman and her colleagues across treatments can be reconciled with the theoretical model presented here. For example, privacy measures require a greater input by the dictator, and less giving, therefore, is consistent with Proposition 1A. Dictators are more self-interested, according to Proposition 2B, if their sensitivity to unfairness is diminished, perhaps as a result of aspects of the experiment that disparage fairness as an issue such as unfair examples in the instructions and the impossibility of fair outcomes built into the structure of

certain treatments. Finally, subjects are less generous, according to Proposition 2C, if the cost of self-deception is lowered, perhaps because of formulations that facilitate the self-deceptive manipulation of accountability and other justice principles—for instance, winning a contest and being told repeatedly by a professor that they have earned the right to allocate.

Reversing the test, it is unclear how the arguments of Hoffman et al. (1996) apply to the results of this paper where social distance is held constant but generosity differs significantly across treatments. It seems more parsimonious to appeal to internalized moral preferences than to erroneous and irrational expectations of future reciprocity. Moreover, if subjects, in fact, stubbornly carry these expectations into the laboratory and act on them even when they are clearly mistaken, has not fairness become internalized in some fashion?¹⁷ James Andreoni and John H. Miller (1998) propose and experimentally test a theory of preferences for giving. Their results are generally consistent with the characterization in this paper of preferences over allocations that involve trade-offs between self- and other-oriented goals.

Many of the successes of economics can probably be attributed to its pushing the assumption of self-interest to the extreme. To proceed further, however, it may be necessary to incorporate richer behavioral assumptions that include fairness and other moral standards. Experimental studies have come to different conclusions about the importance of fairness, but most suggest a nontrivial impact, even though the laboratory context with single-shot decisions and anonymous counterparts probably represents the minimum role for fairness. Still, even a small intrinsic concern for justice, when reinforced by and amplified in social contexts

¹⁷ The source of fairness values remains an unresolved issue, but social context certainly seems to be a reinforcing factor. The suggestion of Hoffman et al. (1998) that people may be preprogrammed or hardwired to accept these norms, as presumed with language acquisition, seems consistent with Mikula's observation (1972a) that children display an increasingly sophisticated sense of justice with age. An apparent difference, however, is that languages vary widely, whereas justice values, based on the limited evidence, seem quite similar across cultures [see, for example, Kahneman et al. (1986b), Bruno S. Frey and Beat Gygi (1988), Roth et al. (1991), and Robert J. Shiller et al. (1991)].

and when distorted by the forces of self-interest, may have significant effects on litigation, wage structure, taxation, regulation, product pricing, and social legislation.

REFERENCES

- Adams, J. Stacy.** "Inequity in Social Exchange," in Leonard Berkowitz, ed., *Advances in experimental social psychology*. New York: Academic Press, 1965, pp. 267–99.
- Akerlof, George A. and Dickens, William T.** "The Economic Consequences of Cognitive Dissonance." *American Economic Review*, June 1982, 72(3), pp. 307–19.
- Andreoni, James and Miller, John H.** "Giving according to GARP: An Experimental Test of the Rationality of Altruism." Mimeo, University of Wisconsin, Madison, November 1998.
- Annenberg/CPB Collection.** "Affirmative Action versus Reverse Discrimination." Program 12 of video series *The Constitution: That delicate balance*, 1984.
- Aronson, Elliot.** *The social animal*, 2nd Ed. San Francisco: W.H. Freeman, 1976.
- Axsom, Danny and Cooper, Joel.** "Reducing Weight by Reducing Dissonance: The Role of Effort Justification in Inducing Weight Loss," in Elliot Aronson, ed., *Readings about the social animal*. San Francisco: W.H. Freeman, 1981, pp. 181–96.
- Babcock, Linda and Loewenstein, George.** "Explaining Bargaining Impasse: The Role of Self-Serving Biases." *Journal of Economic Perspectives*, Winter 1997, 11(1), pp. 109–26.
- Babcock, Linda; Loewenstein, George; Issacharoff, Samuel and Camerer, Colin.** "Biased Judgments of Fairness in Bargaining." *American Economic Review*, December 1995, 85(5), pp. 1337–43.
- Babcock, Linda; Wang, Xianghong and Loewenstein, George.** "Choosing the Wrong Pond: Social Comparisons in Negotiations that Reflect a Self-Serving Bias." *Quarterly Journal of Economics*, February 1996, 111(1), pp. 1–19.
- Blau, Peter M.** *Exchange and power in social life*. New York: Wiley, 1964.
- Bolton, Gary E.** "A Comparative Model of Bargaining: Theory and Evidence." *American Economic Review*, December 1991, 81(5), pp. 1096–136.
- Bolton, Gary E.; Katok, Elena and Zwick, Rami.** "Dictator Game Giving: Rules of Fairness Versus Acts of Kindness." *International Journal of Game Theory*, July 1998, 27(2), pp. 269–99.
- Cason, Timothy N. and Mui, Vai-Lam.** "A Laboratory Study of Group Polarization in the Team Dictator Game." *Economic Journal*, September 1997, 107(444), pp. 1465–83.
- Festinger, Leon.** *A theory of cognitive dissonance*. Stanford, CA: Stanford University Press, 1957.
- Forsythe, Robert; Horowitz, Joel L.; Savin, N. E. and Sefton, Martin.** "Fairness in Simple Bargaining Experiments." *Games and Economic Behavior*, May 1994, 6(3), pp. 347–69.
- Frey, Bruno S. and Gygi, Beat.** "Die Fairness von Preisen." *Schweizerische Zeitschrift für Volkswirtschaft und Statistik*, December 1988, 124(4), pp. 519–41.
- Güth, Werner; Schmittberger, Rolf and Schwarze, Bernd.** "An Experimental Analysis of Ultimatum Bargaining." *Journal of Economic Behavior and Organization*, December 1982, 3(4), pp. 367–88.
- Hoffman, Elizabeth; McCabe, Kevin; Shachat, Keith and Smith, Vernon.** "Preferences, Property Rights, and Anonymity in Bargaining Games." *Games and Economic Behavior*, November 1994, 7(3), pp. 346–80.
- Hoffman, Elizabeth; McCabe, Kevin and Smith, Vernon.** "Social Distance and Other-Regarding Behavior in Dictator Games." *American Economic Review*, June 1996, 86(3), pp. 653–60.
- _____. "Behavioral Foundations of Reciprocity: Experimental Economics and Evolutionary Psychology." *Economic Inquiry*, July 1998, 36(3), pp. 335–52.
- Homans, George C.** "Social Behavior as Exchange." *American Journal of Sociology*, 1958, 63, pp. 597–606.
- Isaac, R. Mark; Mathieu, Deborah and Zajac, Edward E.** "Institutional Framing and Perceptions of Fairness." *Constitutional Political Economy*, Fall 1991, 2(3), pp. 329–70.
- Kahneman, Daniel; Knetsch, Jack L. and Thaler, Richard H.** "Fairness and the Assumptions of Economics." *Journal of Business*, October 1986a, Pt. 2, 59(4), pp. 285–300.
- _____. "Fairness as a Constraint on Profit Seeking Entitlements in the Market." *Ameri-*

- can Economic Review*, September 1986b, 76(4), pp. 728–41.
- Konow, James.** “A Positive Theory of Economic Fairness.” *Journal of Economic Behavior and Organization*, October 1996, 31(1), pp. 13–35.
- _____. “Fair and Square: Four Elements of Distributive Justice.” Presentation to Allied Social Science Meetings, New York, January 1999.
- Lepper, Mark R.; Zanna, Mark P. and Abelson, Robert P.** “Cognitive Irreversibility in a Dissonance-Reduction Situation.” *Journal of Personality and Social Psychology*, October 1970, 16(2), pp. 191–98.
- Loewenstein, George; Issacharoff, Samuel; Camerer, Colin and Babcock, Linda.** “Self-Serving Assessments of Fairness and Pretrial Bargaining.” *Journal of Legal Studies*, January 1993, 22(1), pp. 135–59.
- Los Angeles Times.** “Senate Approves Tax Bill, 74 to 23.” September 28, 1986, pp. A1, A8.
- Messick, David M. and Sentis, Keith.** “Fairness and Preference.” *Journal of Experimental Social Psychology*, July 1979, 15(4), pp. 418–34.
- Mikula, Gerold.** “Die Entwicklung des Gewinnaufteilungsverhaltens bei Kindern und Jugendlichen: Eine Untersuchung an 5-, 7-, 9- und 11jährigen.” *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 1972a, 4(3), pp. 151–64.
- _____. “Gewinnaufteilungsverhalten in Dyaden bei variiertem Leistungsverhältnis.” *Zeitschrift für Sozialpsychologie*, 1972b, 3(2), pp. 126–33.
- Mikula, Gerold and Uray, Hans.** “Die Vernachlässigung individueller Leistungen bei der Lohnaufteilung in Sozialsituationen.” *Zeitschrift für Sozialpsychologie*, 1973, 4(2), pp. 136–44.
- Rabin, Matthew.** “Incorporating Fairness into Game Theory and Economics.” *American Economic Review*, December 1993, 83(5), pp. 1281–302.
- _____. “Cognitive Dissonance and Social Change.” *Journal of Economic Behavior and Organization*, March 1994, 23(2), pp. 177–94.
- _____. “Moral Preferences, Moral Constraints, and Self-Serving Biases.” Working Paper no. 95-241, University of California, Berkeley, August 1995.
- Roth, Alvin E. and Murnighan, J. Keith.** “The Role of Information in Bargaining: An Experimental Study.” *Econometrica*, September 1982, 50(5), pp. 1123–42.
- Roth, Alvin E.; Prasnikar, Vesna; Okuno-Fujiwara, Masahiro and Zamir, Shmuel.** “Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An Experimental Study.” *American Economic Review*, December 1991, 81(5), pp. 1068–95.
- Sefton, Martin.** “Incentives in Simple Bargaining Games.” *Journal of Economic Psychology*, June 1992, 13(2), pp. 263–76.
- Selten, Reinhard.** “The Equity Principle in Economic Behavior,” in H. Gottinger and W. Leinfellner, eds., *Decision theory and social ethics, issues in social choice*. Dordrecht: Reifel Publishing, 1978, pp. 289–301.
- Shiller, Robert J.; Boycko, Maxim and Korobov, Vladimir.** “Popular Attitudes Toward Free Markets: The Soviet Union and the United States Compared.” *American Economic Review*, June 1991, 81(3), pp. 385–400.
- Spiegel, Matthew; Currie, Janet; Sonnenschein, Hugo and Sen, Arunava.** “Understanding When Agents Are Fairmen or Gamesmen.” *Games and Economic Behavior*, July 1994, 7(1), pp. 104–15.
- Walster, Elaine; Walster, G. William and Berscheid, Ellen.** “New Directions in Equity Research.” *Journal of Personality and Social Psychology*, February 1973, 25(2), pp. 151–76.
- Wicklund, Robert A. and Brehm, Jack W.** *Perspectives on cognitive dissonance*. New York: Wiley, 1976.
- Young, H. Peyton.** *Equity in theory and practice*. Princeton, NJ: Princeton University Press, 1994.
- Zajac, Edward E.** *Political economy of fairness*. Cambridge, MA: MIT Press, 1995.
- Zimbardo, Philip G., ed.** *The cognitive control of motivation*. Atlanta, GA: Scott, Foresman, 1969.

