



Penn Institute for Economic Research
Department of Economics
University of Pennsylvania
3718 Locust Walk
Philadelphia, PA 19104-6297
pier@econ.upenn.edu
<http://www.econ.upenn.edu/pier>

PIER Working Paper 09-014

“Ashamed to be Selfish”
Second Version

by

David Dillenberger and Philipp Sadowski

<http://ssrn.com/abstract=1384176>

Ashamed to be Selfish*

David Dillenberger[†] Philipp Sadowski[‡]

April 14, 2009

Abstract

We study a two-stage choice problem. In the first stage, the decision maker (DM) chooses a set of payoff-allocations between herself and a passive recipient. In the second stage, DM chooses an allocation from the set. The recipient is only aware of the second stage choice. Choosing selfishly in the second stage, in the face of a fairer available alternative, may inflict shame on DM. We axiomatize a representation of DM's preferences over sets that identifies DM's selfish ranking, her norm of fairness and shame. It has been suggested that altruism is a prominent motive for non-selfish choice. We identify a condition under which shame to be selfish can mimic altruism, when the experimenter only records the second stage choice. An additional condition implies that the norm of fairness can be characterized as the Nash solution of a bargaining game induced by the second-stage choice problem. The representation is applied to a simple strategic situation, a game of trust.

JEL Classifications: C78, D63, D64, D80, D81

1. Introduction

1.1. Motivation

The notions of fairness and altruism have attracted the attention of economists in different contexts. The relevance of these motives to decision making is intuitive and has been extensively studied. For example, in a classic “dictator game,” where one person gets to anonymously divide, say, \$10 between herself and another person, people tend not to take the whole amount for themselves, but to give a sum between \$0 and \$5 to the other player (for a review, see Camerer (2003)). They act as if they are trading off a concern for fairness

*We thank Roland Benabou and Wolfgang Pesendorfer for their invaluable support. We are also grateful to Eric Maskin, Stephen Morris, Andrew Postlewaite, Charles Roddie and Tymon Tatur for helpful suggestions. This paper was written in part while the authors were graduate prize fellows at the University Center for Human Values, Princeton University. Financial support from the NSF under grant SES-0550540 is gratefully acknowledged.

[†]Department of Economics, University of Pennsylvania. E-mail: ddill@sas.upenn.edu

[‡]Department of Economics, Duke University. E-mail: p.sadowski@duke.edu

or for the other person's incremental wealth and a concern for their own. Thus, preferences for fairness as well as preferences for altruism have been considered (for example, Fehr and Schmidt (1999), Anderoni and Miller (2002), and Charness and Rabin (2002)).

Recent experiments, however, have challenged this interpretation. For example, Dana, Cain and Dawes (2006) study a variant of the same dictator game, where the dictator is given the option to exit the game before the recipient learns it is being played. In case she opts out, she is given a prespecified amount of money and the recipient gets nothing. About a third of the participants choose to leave the game when offered \$9 for themselves and \$0 for the recipient. Write this allocation as (\$9, \$0). Such behavior contradicts altruistic concern regarding the recipient's payoff, because then the allocation (\$9, \$1) should be strictly preferred. It also contradicts purely selfish preferences, as (\$10, \$0) would be preferred to (\$9, \$0). Instead, people seem to suffer from behaving selfishly in a choice situation where they could dictate a fairer allocation. Therefore, they try to avoid getting into such a situation, if they can. Two examples of real-life scenarios would be:

- donating to a charity over the phone, but wishing not to have been home when the call came
- crossing the road to avoid meeting a beggar

We contend that whether or not a person's actions are observed by someone who is affected by her choice plays a crucial role in determining her behavior.¹ We term "shame" the motive that distinguishes choice behavior when observed from choice behavior when not observed. In our model, individuals are selfish when not observed. Thus, concern for another person's payoff is motivated not by altruism, but by avoiding the feeling of shame that comes from behaving selfishly when observed.^{2,3} The interpretation is that, if people are observed, they feel shame when they do not choose the fairest available alternative. Our explanation is supported by further evidence. In a follow-up to the experiment cited above, Dana *et al.* report that only one out of twenty-four dictators exits the game when second-stage choice is also unknown to the recipient. Similarly, Pillutla and Murningham (1995) find evidence

¹We disregard any influence on the player's behavior caused by the presence of the experimenter. In our model, observation by the experimenter is not considered a reason for shame, as the experimenter is not affected by DM's choice.

²To distinguish shame from guilt, note that guilt is typically understood to involve regret, even in private, while, according to Buss (1980), "*shame is essentially public; if no one else knows, there is no basis for shame. [...] Thus, shame does not lead to self-control in private.*" We adopt the interpretation that even observation of a selfish behavior without identification of its purveyor can cause shame.

³Of course, various other-regarding preferences that are not impacted by observation could be present as well (for a comprehensive survey, see Levitt and List (2007)). We do not account for those, as our aim is not to describe a range of possible attitudes toward others, but to highlight shame as a motive for giving. Relaxing our assumptions to allow for some other-regarding preferences, even without observation, would not qualitatively change our results.

that people’s giving behavior under anonymity depends on the information given to the observing recipient. In experiments related to our leading example, Lazear, Malmendier and Weber (2005) as well as Broberg, Ellingsen and Johannesson (2008) predict and find that the most generous dictators are keenest to avoid an environment where they could share with an observing recipient. Broberg *et al.* further elicit the price subjects are willing to pay in order to exit the dictator game; they find that the mean exit reservation price equals 82% of the dictator game endowment. Tadelis (2008) studies a Trust Game and experimentally verifies a probabilistic version of our prediction: When moving from a game with no observation to a game with observation, both the likelihood of cooperation by the receiver and the likelihood of trust by the sender increase.

To better understand the notion of shame and its interaction with selfish preferences, we need to identify the effects of these two motives. A simple and tractable tool for analysis would be a utility that is additively separable in the moral cost (shame) and the private payoff, and that specifies the properties of the shame component (a similar utility is used, for example, by Levitt and List (2007)). We justify using this convenient form by deriving it from plausible assumptions on both preferences and the underlying norm of fairness. To this end, we consider games like the one conceived by Dana *et al.* as a two-stage choice problem. In the first stage, the decision maker (DM) chooses a “menu,” a set of payoff-allocations between herself and the anonymous recipient. This choice is not observed by the recipient. In the second stage, DM chooses an alternative from the menu. This choice is observed, in the sense that the recipient is aware of the menu available to DM.^{4,5} DM has well-defined preferences over sets of alternatives (menus). Our interpretation of shame as the motivating emotion allows considerations of fairness to impact preferences only through their effect on second-stage choices, where the presence of a fairer option reduces the attractiveness of an allocation. Our representation results demonstrate how DM’s norm of fairness and her choice behavior interact. On the one hand, properties of the norm impact choice; on the other hand, the norm of fairness used by DM can be elicited from her choice behavior.

⁴The observed part of the choice procedure is naturally modelled as stage two: The recipient always learns the ultimate choice, as it determines his payoff. Prior to this, DM might be given the option to constrain the set of allocations available for choice. This preceding decision may or may not be observed. If it is not observed, then there is a meaningful first stage. The passage of physical time is not relevant for the distinction of the two stages. This is in contrast to most other models of choice over menus, where subjective uncertainty might be resolved or temptation may kick in over time.

⁵If the exit option is chosen in the aforementioned experiment by Dana *et al.*, as in our setup, the recipient is unaware that there is a dictator who could have chosen another allocation. In their experiment, the recipient is further unaware that another person was involved at all. It would be interesting to see whether informing the recipient that some other person had received \$9 would change the experimental findings. This would correspond to our setup.

1.2. Illustration of Results

Denote a typical menu as $A = \{(a_1, a_2), (b_1, b_2), \dots\}$, where the first and second components of each alternative are, respectively, the private payoff for DM and for the recipient. We impose axioms on DM's preferences over menus that allow us to establish a sequence of representation theorems. To illustrate our results, consider a special case of those representations:

$$U(A) = \max_{(a_1, a_2) \in A} [u(a_1) + \beta \varphi(a_1, a_2)] - \beta \max_{(b_1, b_2) \in A} [\varphi(b_1, b_2)], \quad (*)$$

where u and φ are increasing in all arguments. u is a utility function over private payoffs and $\varphi(a_1, a_2)$ is interpreted as the fairness of the allocation (a_1, a_2) .

Alternatively, if we denote by a^* and b^* the two maximizers above, it can be written as:

$$U(A) = \underbrace{u(a_1^*)}_{\text{value of private payoff}} - \underbrace{\beta(\varphi(b_1^*, b_2^*) - \varphi(a_1^*, a_2^*))}_{\text{shame}}.$$

This representation captures the tension between the impulse to maximize private payoff and the desire to minimize shame from not choosing the fairest alternative within a set. It evaluates a menu by the highest utility an allocation on the menu gets, where this utility depends on the menu itself. The utility function that is used to evaluate allocations is additive and has two distinct components. The first component, $u(a_1)$, gives the value of a degenerate menu (a singleton set) that contains the allocation under consideration. When evaluating degenerate menus, which leave DM with a trivial choice under observation, we assume her to be *Selfish*: she prefers one allocation to another if and only if the former gives her a greater private payoff, independent of the recipient's payoff. The second component is "shame." It represents the cost DM incurs when selecting (a_1, a_2) in the face of the fairest available alternative, (b_1^*, b_2^*) .

As shame is evoked whenever this fairest available alternative is not chosen, we can relate choice to a second, induced binary relation "fairer than", which represents DM's private norm of fairness. Based on the definition that "*fair implies an elimination of one's own feelings, prejudices, and desires so as to achieve a proper balance of conflicting interests.*" (Merriam-Webster Collegiate Dictionary (Tenth Edition, 2001)), DM's private norm of fairness is assumed to satisfy at least the following three properties: *Fairness Ranking*, which implies that the fairness comparison of any two alternatives is independent of the other available options; the *Pareto* criterion on payoffs; and *Compensation*, that allows any variation in the level of one person's payoff to be compensated by appropriate variation in the level of the other person's payoff.

In the special case considered here, the shame from choosing (a_1, a_2) in stage two is

$\beta (\varphi (b_1^*, b_2^*) - \varphi (a_1, a_2))$. This implies that even alternatives that are not chosen may matter for the value of a set, and larger sets are not necessarily better. To see this, consider the representation (*) with $u(a_1) = a_1$, $\beta = \frac{1}{2}$ and $\varphi(a_1, a_2) = a_1 a_2$ and compare the sets $\{(10, 1), (4, 3)\}$, $\{(10, 1)\}$ and $\{(4, 3)\}$. Evaluating these sets we find $U\{(10, 1), (4, 3)\} = 9$, $U\{(10, 1)\} = 10$ and $U\{(4, 3)\} = 4$. To permit such a ranking, we assume a version of *Left Betweenness*, which allows smaller sets to be preferred over larger sets. Theorem 1 establishes that our weakest representation, which captures the intuition discussed thus far, is equivalent to the collection of all the above assumptions.

Representations similar to (*) have been extensively studied in the literature on temptation, starting with the work of Gul and Pesendorfer (2001, henceforth GP). GP consider preferences over menus of lotteries and, furthermore, impose a version of the independence axiom. Whereas we feel that introducing (for technical reasons) uncertainty to an otherwise riskfree environment is debatable, imposing the independence axiom would be simply inappropriate in our context. For example, suppose that the fairness ranking of alternatives is symmetric. Then (9, 1) is as fair as (1, 9). Independence implies that any randomization over these two outcomes is as fair as either of them, while common sense suggests that awarding \$9 to either player with probability $\frac{1}{2}$ would be fairer.⁶ In addition, the independence axiom implies a menu-independent second-stage choice criterion. Contrary to this, we argue that a higher degree of shame may well lead to a fairer choice. Our most general representation, Theorem 1, accommodates such a context-dependent choice criterion.⁷

The special case considered in (*), on the other hand, does feature a context-independent choice criterion. To see this, regroup the terms as follows:

$$U(A) = \underbrace{\max_{(a_1, a_2) \in A} [u(a_1) + \beta \varphi(a_1, a_2)]}_{\text{second stage choice criterion}} - \beta \underbrace{\max_{(b_1, b_2) \in A} [\varphi(b_1, b_2)]}_{\text{effect of fairest alternative}}.$$

Theorem 2 establishes that, given the assumptions made so far, an additional separability assumption on preferences over sets, *Consistency*, is equivalent to the existence of such a choice criterion. Suppose that only the second stage of the choice procedure is observed (for example, because DM, as in the classic dictator game, never gets to choose between menus). If second-stage choice is context independent, shame might be mistaken for altruism: DM

⁶See section 5 for a context that naturally involves uncertainty and our suggestion for incorporating it into our model.

⁷In the context of temptation, Noor and Takeoka (2008) suggest relaxations of the independence axiom that allow menu-dependent choice. In Epstein and Kopylov (2007), the choice objects are menus of acts. They relax independence and characterize a functional form with a convex temptation utility. Independently of our work, Olszewski (2008) studies preferences over subsets of a finite set of deterministic outcomes and finds a representation where both choice and temptation are context dependent.

seems to trade off a selfish concern for his private payoff with a concern for the recipient's welfare. We argue, however, that it is hard to reconcile such an interpretation with any choice reversal in stage two. Thus, when observing stage two in isolation, shame can mimic altruism only if the induced choice ranking is context independent, or equivalently if the ranking of menus satisfies Consistency.

We further specify the norm of fairness by assuming that the private payoffs to the two players have *Independent Fairness Contributions*: Fairness should be concerned with utilities, not monetary payoffs, but interpersonal comparisons of utilities are infeasible. Thus, the fairness contribution of raising one player's monetary payoff can not depend on the level of the other player's payoff. With this additional assumption, Theorem 3 establishes that there are two utility functions, v_1 and v_2 , evaluated in the payoff to DM and the recipient respectively, such that the value of their product represents the fairness ranking, $\varphi(a_1, a_2) = v_1(a_1)v_2(a_2)$. Thus, the fairest alternative within a set of alternatives can be characterized as the Nash Bargaining Solution (NBS) of an associated game. Because the utility functions used to generate this game are private, so is the norm. We argue that when based on true selfish utilities, the NBS is a convincing fairness criterion in our context. Those utilities may not be known to DM (especially in anonymous choice situations) but one can assess the descriptive appeal of the representation by asking whether the utilities comprising DM's norm at least resemble selfish utilities.

Example: Let $u(a_1) = a_1$, $\varphi(a_1, a_2) = v_1(a_1)v_2(a_2) = a_1a_2$ and $\beta = \frac{1}{2}$. This implies that the utilities v , which are used to generate the fairness ranking, coincide with u . Shame is half the difference between the Nash-product of the fairest and the chosen alternatives. In the experiment by Dana *et al.* mentioned above, only whole dollar amounts are possible allocations. The set $A = \{(10, 0) (9, 1) (8, 2), \dots, (0, 10)\}$ then describes the dictator game. It induces the imaginary bargaining game with possible utility-allocations $\{(10, 0), (9, 1), (8, 2), \dots, (0, 10), (0, 0)\}$, where the imaginary disagreement point is $\lim_{(x,y) \rightarrow 0} (v_1^{-1}(x), v_2^{-1}(y)) = (0, 0)$. According to the NBS, $(5, 5)$ would be the outcome of the bargaining game. Its fairness is $5 \cdot 5 = 25$. To trade off shame with selfishness, DM chooses the alternative that maximizes the sum of private utility and fairness, $a_1 + a_1a_2$, which is $(6, 4)$. Its fairness is $6 \cdot 4 = 24$ and the shame incurred by choosing it is $\frac{1}{2}$. Hence $U(A) = 5.5$. From the singleton set $B = \{(9, 0)\}$, which corresponds to the exit option in the experiment, the choice is trivial and $U(B) = 9$. This example illustrates both the trade-off DM faces when choosing from a non-degenerate menu and the reason why she might prefer a smaller menu.

The organization of the paper is as follows: Section 2 presents the basic model and a representation that captures the concepts of fairness and shame. Section 3 isolates a choice criterion from the choice situation. Section 4 further specifies the fairness ranking. Section 5 suggests an application to a simple strategic situation, a game of trust. Section 6 concludes by pointing out connections to existing literature. An extension of our results to incorporate multiple recipients and all proofs are relegated to the appendix.

2. The Model

Let K be the set of all finite subsets of \mathbb{R}_+^2 .⁸ Any element $A \in K$ is a finite set of alternatives. A typical alternative $\mathbf{a} = (a_1, a_2)$ is interpreted as a payoff pair, where a_1 is the private payoff for DM, and a_2 is the private payoff allocated to the (potentially anonymous) other player, the recipient.⁹ Endow K with the topology generated by the Hausdorff metric, which is defined for any pair of non-empty sets, $A, B \in K$, by:

$$d_h(A, B) := \max \left[\max_{\mathbf{a} \in A} \min_{\mathbf{b} \in B} d(\mathbf{a}, \mathbf{b}), \max_{\mathbf{b} \in B} \min_{\mathbf{a} \in A} d(\mathbf{a}, \mathbf{b}) \right],$$

where $d : \mathbb{R}_+^2 \rightarrow \mathbb{R}_+$ is the standard Euclidian distance.

Let \succ be a continuous, strict preference relation over K . The associated weak preference, \succeq and the indifference relation, \sim are defined in the usual way.

The choice of a menu $A \in K$ is not observed by the recipient, while the choice from any menu is. We call the impact this observation has on choice "shame." The first axiom specifies DM's preferences over singleton sets.

P_1 (Selfishness) $\{\mathbf{a}\} \succ \{\mathbf{b}\}$ if and only if $a_1 > b_1$.

A singleton set $\{\mathbf{a}\}$ is a degenerate menu that contains only one feasible allocation, (a_1, a_2) . It leaves DM with a trivial choice to be made when being observed in the second stage. Therefore, the ranking over singleton sets can be thought of as the ranking over allocations that are imposed on DM. We contend that there is no room for shame in this situation; choosing between two singleton sets reveals DM's "true" preferences over allocation outcomes. The axiom states that DM is not concerned about the payoff to the second player when evaluating such sets; she compares any pair of alternatives based solely on the first component, her private payoff. If, for example, DM had an altruistic concern for fairness

⁸With \mathbb{R}_+ we denote the positive reals including 0. \mathbb{R}_{++} denotes the positive reals without 0.

⁹The extension to the case where K is the set of all finite subsets of \mathbb{R}_+^n , that is, the case where DM is concerned about the welfare of other $n-1$ recipients, is given in the appendix.

