Trust, Truth, Status and Identity
an experimental inquiry

Jeffrey V. Butler[*]
Dept of Economics, University of California, Berkeley
butlerj@econ.berkeley.edu

Abstract: I report the results of an experiment demonstrating that such norm-based behaviors as trust, reciprocity and truth-telling vary with *social* identity in predictable and potentially economically important ways. There were two versions of the experiment—one version in which two equal-status social identities were induced, and a second version where two unequal-status social identities were induced. In both versions, subjects played two standard economic games: the Trust Game, as well as a two-player asymmetric information game.

Comparing subjects' behavior across the two versions of the experiment allows for a succinct description of the effects of social identity: subjects held members of their own group to a higher standard; and high status subjects held everyone to a higher standard. This latter effect—"high status/high standards"—was clearest in the Trust Game: high status subjects both trusted more and punished co-players' lack of trust more severely than their low status counterparts. To make the high status/high standards hypothesis more concrete, I constructed a simple identity model and used a semi-parametric technique to estimate what subjects' "standards" were.

[*] I owe a great debt of gratitude to George A. Akerlof, Shachar Kariv, John Morgan and Steven Marc Goldman for untold hours of discussion and many helpful comments. I am also grateful to to the National Science Foundation and UC Berkeley's Xlab for financial support.

1 Introduction

An ever-growing body of experimental research in economics demonstrates undeniable and systematic deviations from pure self-interest. Many of the situations in which these deviations have been documented share a particularly simple form: what a purely self-interested person would do is diametrically opposed to what a "decent" person *should* do. For lack of a better term, call what one *should* do a norm. And call the aforementioned class of results norm-based deviations from pure self-interest.

The simplest example of this type of behavior involves a sharing norm: we've all been taught since grade school that we *should* share. Consequently, there is considerable experimental evidence that subjects share money earnings in the lab even when there is no plausible purely self-interested reason for doing so. Beyond simply sharing, however, the class of norm-based deviations from pure self-interest encompasses phenomena such as trust, reciprocity and altruism which are thought to be central to everything from the functioning of labor markets (Bewley, 1999; Akerlof, 1982) to aggregate economic growth (Knack and Keefer, 1997). For an overview of many economics experiments dealing with these phenomena, see Camerer (2003).

In terms of economic theory, there are two main approaches to explaining norm-based deviations from pure self-interest. In the "social identity" approach—represented in the economics literature by Akerlof and Kranton (2005)—the building blocks are categorizations. Individuals place themselves and others into social categories. Each category is a social identity. Norms in this view—how we *should* and *should not* behave—are tied inextricably to social identities. To complete the setup, individuals have preferences over their own and others' norm-compliance.

In another, more established, approach, norms are modeled as stable, *individual* traits. In these models—generally referred to as "social preferences" models—norm-concern can be boiled down to individual-specific parameters, with norm-based variation in behavior explained by individual heterogeneity in these parameters. Sometimes, informational phenomena such as signaling complement the explanations. (See, e.g., Benabou and Tirole, 2004; Levine, 1998; and Charness and Rabin, 2002.) Some of the most widely-cited social preferences models make the further assumption that norms are the same for everyone—i.e., assuming that one norm is *the* norm. Behavioral heterogeneity in these models stems from differences in how much individuals care about *the* norm vis-a-vis standard economic incentives such as wealth preferences. Again, this tradeoff is modeled as a stable, individual-specific trait: some people are purely self-interested, some people care only about, e.g., "fairness," and most people lie somewhere in between (e.g., Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000).

The key difference between the social identity and social preferences approaches is where norms reside, and this difference leads to a testable hypothesis. Specifically, since the social identity approach asserts that norms are tied to social categories, manipulating categories will change observed behavior in situations where norms are important. On the other hand, if we are careful to not change anything systematically about the *individuals* involved, the social *preferences* approach would predict that merely changing categorizations will not produce significant changes in behavior.

Herein I report the results of an experiment in which I tested this hypothesis by varying subjects' social identities in a manner that did not change anything typically considered decision-relevant about the individuals involved: social identities were assigned randomly. The results support the social identity approach in that even these simple, information-less categorizations affected subjects' behavior. Loosely speaking, subjects held members of their own social identity to a higher standard, and, when status differences were introduced between groups, this in-group bias disappeared. In the latter situation, high status subjects held *everyone* to a higher standard.

Before I can describe the results more precisely, it will prove helpful to give an example what I mean by "social identity." The concept of social identity is vividly illustrated by a classic experiment: Robbers Cave (Sherif, *et al*, 1954.) In the Robbers Cave experiment, two groups of otherwise-similar schoolboys were sequestered in separate camps at Robbers Cave State Park in Oklahoma. During this initial week-long phase, each group engaged in cohesion-building activities: running, hiking, swimming—standard summer camp fare. The groups became cohesive enough to spontaneously name themselves: one group called itself the Rattlers, while the other group deemed itself the Eagles.

In the second phase, the two groups engaged in competitive intergroup activities—baseball tournaments, tug-of-war and the like. To put it dryly, in the second phase each group demonstrated out-group aversion. The groups' aversion to each other was sufficiently strong to induce minor acts of arson and multiple attempted assaults. For example, one attempted assault stemmed from the Rattlers' desire to settle the score for a midnight raid on their cabin by the Eagles, and featured a mess hall raid where Rattlers armed themselves with sticks and bats. A later attempt involved the Rattlers lying in wait for the Eagles with rock-filled socks as weapons.

In a social identity framework, we can think of the experiment as creating two social identities—Eagles and Rattlers. And, the experiment revealed the existence of a particular (social) identity-based norm—out-group aversion. The strength of the manipulation obviously raises many concerns about whether this was a purely identity-based phenomenon. Most of the obvious alternative explanations have been ruled out by four decades of experimental work in social psychology, starting with Tajfel, *et al* (1971). In this vein of experimental research, out-group aversion has been shown to be a significant phenomenon under much more pallid, controlled circumstances—including simply randomizing experimental subjects into two different groups.

Given this background, the current experiment proceeded in two phases: an identity-inducement phase and a game-playing phase. Following the social psychology literature, the identity-inducement phase created two social identities by randomly dividing subjects into two groups. In one version of the experiment, the two groups were of equal status. In a second version of the experiment—following, in the economics literature, Eckel, Ball, Grossman and Zame (2001)—I randomly chose one of the groups to have "high status," and reinforced subjects' sense of status. Here, status was reinforced by varying tasks and comfort levels: high status subjects were seated three per row while low status subjects sat in rows of five; and high status subjects enjoyed refreshments while low status subjects performed a boring, tedious task: re-alphabetizing a list of names by hand. These status-reinforcement activities lasted for ten minutes. I will provide evidence later in the paper suggesting that the status manipulation did not simply induce a mood effect or a wealth effect.

The game-playing phase of the experiment was standard experimental economics fare. Subjects played two widely-researched economic games. In the Trust Game (Berg, Dickhaut and McCabe, 1995), one subject—the sender—is given a fixed sum of money, of which he can send some, all or none of to his anonymous co-player—the receiver. The amount sent is tripled, at which point the receiver can return any of this tripled amount to the sender. The sender's action involves trust, as his co-player has the option of returning nothing; and the receiver's action involves reciprocity, as pure self-interest dictates keeping the entire amount sent, while rewarding "nice" actions requires returning a non-zero amount.

The second game was a two-player asymmetric information game I call the Truth Game. The Truth Game can be thought of as modeling the interaction between a used car salesman and a potential customer. The salesman has private information about the quality of the car, while the buyer must decide to either buy the car or walk away. The salesman, of course, can claim the car is reliable, but the buyer has no way to verify the salesman's claims. The gist of the strategic situation is that salesmen can lie to buyers for potential monetary gain.

The relevance of these games to the current inquiry is that, in both of these games, normative behavior (trust and honesty, respectively) is at odds with purely self-interested behavior. Additionally, these two games represent situations that many economists consider important.[2] Finally, while the Trust Game has a long history in the social preferences literature, results in the Truth Game will demonstrate the predictive power gained from a social identity approach.

There were three main results in the data. Firstly, I found that equal-status identities induced a form of in-group bias: subjects held members of their own group to a higher standard. Secondly, introducing unequal status eliminated this form of in-group bias and replaced it with a pattern that can be succinctly summarized: high status subjects held everyone—including themselves—to higher standards. Specifically, high status subjects were both more trusting in the Trust Game and more honest in the Truth Game; and in the Trust Game, where there is an opportunity to punish "bad" behavior, high status subjects both punished lack of trust more severely, and rewarded high levels of trust more generously, than their low status counterparts. Thirdly, across-version comparisons of reciprocal behavior in the Trust Game yielded insights into why status affects behavior: the data suggest that high status emboldened subjects to impose their values on everyone in precisely the same manner that subjects were emboldened to impose their values on their "own" people *sans* status differences. This effect, too, can be summarized succinctly: "give someone an office and they become officious."[3]

Additionally, across-version comparisons of behavior in the Truth Game revealed a potentially important consequence of status differences: contrary to all purely self-interested equilibria of the Truth Game, in the unequal-status version of the experiment, sellers' messages actually benefitted buyers. In contrast, in the equal-status version of the experiment, buyers did no better than if they had ignored sellers' messages altogether.

[2] The standard quote concerning the importance of trust, and hence the Trust Game, is due to Arrow (1972): "Virtually every economic transaction has within it an element of trust . . . ." The strategic situation in the Truth Game is clearly at the heart of many situations involving asymmetric information: contracting with an expert, voting based on campaign promises, etc.

[3] The source of this expression is my father-in-law, who gleaned this phenomenon from decades of counseling individuals and dealing with large organizations as the senior pastor of several large churches throughout the U.S.

To investigate the high status/high standards phenomena further, following Akerlof and Kranton (2005), I constructed a simple social identity model of subjects' preferences having the following form, where a_k denotes player $k's$ action:

$$U_j = u(a_j, a_i) - \alpha_j(a_j - a^{Ideal}_{c_j,c_i})^2 \tag{1}$$

In this two-player social identity model, player j derives utility from two sources: the purely self-interested preferences of classical economics—$u(a_j, a_i)$—and from living up to her *ideals*—$a^{Ideal}_{c_j,c_i}$. An ideal is the specific action prescribed by relevant social norms, and depends on the social identities of both players—c_j and c_i in Equation 1.[4] To make the distinction clear, the relevant *norm* in the Trust Game is trust; while one possible *ideal* is "send exactly seven dollars to members of your own social identity, otherwise send nothing." Finally, player j is willing to trade economic utility against "identity utility" at a rate measured by α_j.

Given this model, I used a semi-parametric estimation technique—Censored Least Absolute Deviations (Powell, 1984)—to estimate subjects' ideals explicitly. The estimates suggest that high status subjects punished norm-deviance, ideally, more than twice as severely as low status subjects; on the other hand, when co-players fully complied with norms, high status subjects were more generous than low status subjects.

The remainder of this paper is organized as follows. First, I provide a brief sketch of related literature. Next, the experimental design is presented in detail. After this, results are presented for each game, separately. Next, I construct a simple identity model and use receivers' actions in the Trust Game to estimate receivers' Ideals. Subsequently, I provide some evidence that the results cannot be plausibly explained by either mood effects or wealth effects. In the concluding section, I summarize the results and suggest future avenues of research.

2 Related Literature

The identity inducement phase of the current experiment draws mainly on the body of literature in social psychology referred to as the "minimal group paradigm" (MGP). In MGP experiments, investigators randomly divide subjects into two groups; group affiliation constitutes a social identity. Each subject then divides a sum of money between two *other* subjects, knowing

[4] The "c" refers to social **c**ategory; using "i" for social **i**dentity would be more natural, but also more confusing in terms of notation.

	Money Initially Sent			
	5	10	15	20
Avg Money Returned				
by Ashkenazic:	1.8	13	17.2	24.3
by Eastern:	2.8	14.2	16.7	23.1

Table 1: Trust Game (Fershtman and Gneezy, 2001)

only the social identity of these two other subjects. A robust finding of hundreds of MGP experiments is that subjects allocate significantly more money to members of their own group—i.e., out-group aversion, also known as in-group bias. For a meta-analysis of dozens of MGP experiments, see Mullen, Brown and Smith (1992). For a more recent overview of this research, see Haslam (2001). And for a more critical survey, see Brown (2000).

Of course, the MGP experiments are too cute by half. In particular, self-interest is not involved since the money-division tasks are disinterested. This leaves open the possibility that social identity effects are sufficiently weak to be drowned out whenever self-interest *is* implicated—in which case, social identity could be safely ignored by economists.

There are clues in the existing experimental economics literature that social identity affects behavior when self-interest *is* implicated. Two of the clearest such clues appear in Glaeser, Laibson, Scheinkman and Soutter (2000) and Fershtman and Gneezy (2001)—hereafter referred to as GLSS and FG, respectively. Both experiments involved mostly-standard Trust Games. Not coincidentally, GLSS provides clues about the effects of two equal-status identities; while FG provides clues concerning the effects of two unequal-status identities.[5]

In FG, the subjects were Ashkenazic and "Eastern" Jews. Ashkenazic Jews can be thought of as having higher status, as they "achieve higher levels of education and earnings than do Eastern immigrants. . . "[15, p. 352]. An unexplained pattern in the data was that Ashkenazic Jews (high status) both trusted and reciprocated trust more than Eastern Jews (low status). The reciprocity patterns are clear in Table 1 below (reproduced from Table 2 (pg 363) in FG): when sent a low amount, Ashkenazic Jews return less than Eastern Jews; however, when sent a high amount, they return *more* than Eastern Jews.

[5] I say "clues," rather than evidence, because neither of these experiments had the primary purpose of investigating social identity, and therefore have experimental designs which cannot rule out compelling alternative explanations.

In GLSS, two groups of subjects involved were Asian students and white students. Being drawn from the same (Harvard) economics class, these groups likely have equal status. One pattern in the data was that subjects were less trustworthy when paired with a subject of different ethnicity. Importantly, the pattern was not "members of certain ethnic groups are less trustworthy," but rather, the implied pattern was that the *same* person could be more or less trustworthy depending on the ethnicity of their co-player.

As intriguing as these clues are, there is an obvious confound in experiments: group affiliations convey possibly decision-relevant information. Specifically, the informational confound in GLSS stems from the fact that subject pairings were *not anonymous*. Additionally, patterns of association may make a person of the same ethnicity more likely to share experimental earnings outside of the experiment. In FG, group affiliation clearly conveyed information about future earnings, which could have been a factor in second-movers' decisions.

The current experiment recreates the essence of the social identities in FG and GLSS, while controlling for the informational confound. Following the minimal group paradigm, I randomly assigned subjects to one of two groups. In one version of the experiment the two groups were of equal status (GLSS); in a second version of the experiment, I randomly chose one of the groups to have "high status" (FG).

Another closely related paper in experimental economics is Ball, Eckel, Grossman and Zame (2001). There, the authors randomly assigned status levels and conducted a two-sided auction, finding that prices tended to favor high status subjects: goods traded at higher prices when sellers had high status and lower prices when *buyers* had high status. This is consistent with the high status/high standards hypothesis—it could result, for instance, if buying high and selling low was the norm. However, the authors explain their results with the assumption that agents prefer to interact with higher status agents, and the design of the experiment does not allow one to distinguish between these two explanations.

In the social psychology literature, variants of the high status/high standards phenomenons can be found as far back as Homans' analysis (1950, p. 141) of a famous and even earlier quasi-experiment, the Bank Wiring Observation Room (details appear, among other places, in Roethliserger and Dickinson (1939).) Many of the minimal group experiments involving status are also consistent with such an hypothesis. See, e.g., any of the overviews mentioned above, or, for the seminal work, see Turner and Brown (1978).

3 Experimental Design

The subjects in the experiments were recruited from undergraduates and staff at the University of California, Berkeley. During the course of the experiment, each subject played from ten to fifteen rounds of a standard Trust Game as well as ten to fifteen rounds of a costless signalling game. The experiment was programmed and conducted with the software z-Tree (Fischbacher 2007), and all sessions were conducted in the X-lab facilities at the University of California, Berkeley. All together, eight sessions were conducted and 144 subjects participated.[6]

The experiment consisted of two phases: an initial identity-inducement phase, followed by a game-playing phase. There were two possible versions of the identity inducement phase: the identity-only version (ID-only), and the status and identity version (S-ID). The game-playing phase was identical in each version of the experiment. Each subject participated in only one of the two possible versions.

3.1 Identity Inducement Phase

In the ID-only version of the experiment, subjects were randomly assigned one of two colors—purple or orange—and then seated on one side of the room or the other, according to color group. Randomization was achieved by publicly drawing poker chips out of a canvas bag, and it was made clear that purple and orange were equally likely to be drawn. Each color can be thought of as an identity.

Each color group's seating arrangement was the same, and the groups faced each other across empty space in the middle of the room. After they were seated, I handed each subject a wristband matching their assigned color, which they were instructed to wear for the duration of the experiment. After this, subjects proceeded directly to the game-playing phase.

In the S-ID version, subjects were randomly divided into color groups in exactly the same manner as in the ID-only version. And subjects were, again, given a wristband corresponding to their color group and asked to wear it during the experiment. The difference between the S-ID and the

[6] There were actually 9 sessions conducted, involving 160 subjects. One session experienced such severe technical glitches that the data could not be used. Specifically, the glitch required me to unnecessarily make myself a focal point of the experiment—as suggested by comments on the post-experiment questionnaire—and to dismiss two subjects. Further complicating this session was the fact that both subjects I dismissed were in the same status category which could have possibly introduced confounding majority/minorty effects.

ID-only version was that in the S-ID version, one of the colors groups was publicly randomly chosen to signify “high status,” while the non-chosen color signified “low status.”

Status was assigned by putting one purple poker chip and one orange poker chip in a canvas bag, shaking the bag in full view of all subjects, then drawing one poker chip out of the bag. The drawn poker chip was held up by the experimenter, after which it was announced that the color of the poker chip drawn out of the bag would represent “high status” for the duration of the experiment. Status differences were reinforced by varying groups’ comfort levels and assigning status-specific tasks.

To make high status subjects more comfortable than low status subjects, high status subjects were seated three per row, while low status subjects were seated five per row. In terms of tasks, high status subjects were allowed to enjoy refreshments while low status subjects worked. Specifically, low status subjects were assigned a boring and tedious task—taking a list of names alphabetized by last name, and re-alphabetizing the list by first name, by hand. This phase lasted 10 minutes, after which materials were collected and subjects proceeded directly to the game-playing phase.

Out of a total of eight experimental sessions conducted, three of these sessions were ID-only and five sessions were the S-ID version. Overall, 60 subjects participated in the ID-only version of the experiment, while 84 subjects participated in the S-ID version.

3.2 Game-Playing Phase

The game-playing phase was identical in both versions of the experiment: all subjects played multiple rounds of the Trust Game and multiple rounds of a two-player costless signalling game. The Trust Game is a two-player sequential game of perfect information. In this game, first-movers are called “senders” while second-movers are called “receivers.” The game begins with the sender deciding to send some, all or none of his or her endowment—here, $7—to their anonymous co-player. Each dollar sent is then tripled by the experimenter, and allocated to the receiver. Finally, the receiver then decides how much of this tripled amount to send back. Any money not sent back is kept by the receiver. All purely self-interested subgame perfect equilibria of the Trust Game involve the receiver returning nothing, and hence, the sender sending nothing as well.

The costless signalling game—hereafter, the Truth Game—can be thought of as modeling the situation faced by a used car salesman and a prospective buyer. The salesman has private information—whether the car is reliable or

	Buyer Action	
	Buy	Walk Away
Actual Quality		
Reliable	$(12, 12)$	$(10, 10)$
Lemon	$(12, 10)$	$(10, 12)$
(a,b) = ($ Seller, $ Buyer)		

Table 2: Truth Game Payoffs

a lemon. After observing this information, the salesman can send one of two messages to the buyer: he can either tell the buyer that the car is reliable, or that it's a lemon. The buyer observes the seller's message and then can take one of two actions—buy or walk away. The buyer prefers buying the car only if it's reliable, otherwise walking away is the buyer's preferred action. The salesman, on the other hand, prefers the buyer to buy irrespective of the car's quality. Payoffs in the Truth Game depend solely on the quality of the car and the buyer's decision. Monetary payoffs are given in Table 2 and the game tree can be found in the appendix. One feature of the Truth Game is that it has no purely self-interested equilibria in which sellers' messages convey any information. Subjects' instructions for the Truth Game, as well as the Trust Game, can be found in the appendix.

It is worth mentioning that the payoffs in the Truth Game (Table 2) would not change qualitatively if they were to be transformed into utilities in a reasonable social-preferences model. For instance, when the car is a lemon, the buyer is choosing between payoff distributions in which either the buyer is ahead or the seller is ahead. One might reasonably expect that the buyer prefers the former and the seller prefers the latter. Similarly, when the car is reliable, the buyer is choosing between two distributions—one of which dominates the other. A reasonable social preferences model would imply that both players prefer the dominant allocation.

In total there are 760 observations for each game in the data. There are 300 observations for each game stemming from the ID-only version of the experiment, and 460 observations for each game come from the S-ID version of the experiment. Since these totals reflect multiple rounds of game-play, steps were taken to minimize any possible dynamic effects. After each round of each game, subjects were randomly (and anonymously) re-paired and roles within each game were randomly reassigned. For example, a particular subject could be a buyer in the Truth Game one round and a seller the next, and be paired with a high status co-player in one round and a low status co-player in the next round. Furthermore, subjects were never informed of the outcomes of any of the rounds of either game. These steps appear to have been successful: overall, subjects' actions in early rounds of each game did not differ significantly from their actions in later rounds. Therefore, all reported results incorporate aggregating over all rounds within each game.

4 Results

4.1 The Trust Game

4.1.1 Summary Statistics

As in previous research on the Trust Game, senders exhibited a substantial amount of trust. Similar to the original Trust Game experiment (Berg, Dickhaut and McCabe, 1995), senders sent about half the maximum possible amount. While the average amount sent was similar across versions of the experiment, a Kolmogorov-Smirnov test suggests that the *distributions* of amounts sent differed significantly across versions ($p = 0.007$). The key difference between the distributions of amounts sent is the propensity to send all or nothing. Introducing status seems to have swayed fence-sitters, making both complete trust and complete lack of trust more likely at the expense of middling trust levels. A summary of the distributions is presented in Table 3.

For their part, receivers returned close to 80% of the amount senders' sent—excluding observations where senders initially sent zero.[7] Again, this is in line with receivers' decisions in Berg, Dickhaut and McCabe (1995), where the average proportion returned was 89.5%.

Although a Kolmogorov-Smirnov test suggests that the distributions of proportions returned varied across versions ($p = 0.058$), this fact does not have much meaning by itself as it could be an artifact of variation in senders' actions across versions. One way to get a sense of whether, and how, receivers actions varied between the ID-only and S-ID versions of the experiment is to divide the data into two categories—one in which the amount initially sent was above the median, and another in which it was below the median. Using these categories, we see that when faced with a low amount sent, the most generous receivers—i.e., the top 10%—in the ID-only version were more generous than the most generous receivers in the S-ID version.

[7] Since money sent was tripled, the proportion returned can take values from 0 to 3.

Proportion of Maximum Sent		
	ID-only	S-ID
Average Proportion Sent	0.51	0.55
Standard Error	(0.022)	(0.019)
10th percentile	0	0
20th percentile	0.07	0
30th percentile	0.21	0.14
40th percentile	0.36	0.43
50th percentile	0.50	0.57
60th percentile	0.71	0.86
70th percentile	0.86	1
80th percentile	1	1
90th percentile	1	1
N	300	460

1. Maximum possible was \$7, so reported value is $\frac{AmountSent}{7}$.

Table 3: Trust Game Senders, proportions sent

On the other hand, when faced with a high amount sent, receivers in the S-ID version were more likely to display high levels of generosity than receivers in the ID-only version.

Kolmogorov-Smirnov tests fail to reject the null hypothesis of equal distributions when initial amounts sent were less than the median; while, for observations where the initial amount sent was above the median, the distributions were significantly different ($p = 0.009$). A summary of the distributions of proportions returned by receivers is presented in Table 4.

4.1.2 Patterns in Senders' Actions

In both versions of the experiment, the evidence of in-group bias on the part of first-movers was weak, but consistent. While each group sent more, on average, to members of their own group than to members of the other group, the patterns were not statistically significant.

This is not to say there were no patterns, however. Pronounced patterns in senders' actions are present in the S-ID version, where unequal status was introduced. Consistent with the hypothesis that high status senders in the

	Proportion Returned					
	Overall		<= Median Sent		> Median Sent	
	ID-only	S-ID	ID-only	S-ID	ID-only	S-ID
Average Return Proportion	0.77	0.78	0.55	0.54	0.92	0.92
Standard Error	(0.045)	(0.039)	(0.075)	(0.064)	(0.054)	(0.048)
10th percentile	0.00	0.00	0.00	0.00	0.00	0.00
20th percentile	0.00	0.00	0.00	0.00	0.14	0.00
30th percentile	0.14	0.00	0.00	0.00	0.62	0.29
40th percentile	0.43	0.33	0.00	0.00	1.00	0.71
50th percentile	0.75	0.67	0.25	0.25	1.00	1.00
60th percentile	1.00	1.00	0.43	0.50	1.14	1.29
70th percentile	1.14	1.30	0.67	0.83	1.29	1.43
80th percentile	1.33	1.43	1.00	1.00	1.42	1.50
90th percentile	1.50	1.50	1.90	1.50	1.50	1.50
N	249	355	102	131	147	224

1. Since money sent was tripled, the return proportion can take values from 0 to 3.

2. ID-only median sent = 3.50; S-ID median sent = 4.00.

3. Includes only observations where money initially sent was greater than zero.

Table 4: Trust Game Receivers, proportion returned

Trust Game held themselves to higher standards, high status senders trusted their co-players significantly more on average. Overall, high status senders were significantly more trusting than low status senders ($p < 0.01$); and significantly more trusting than their ID-only counterparts. This "extra" trust was extended to both high and low status receivers, as high status senders sent more to both types of S-ID receivers than their low status counterparts. These patterns are evident in Table 5.

4.1.3 Patterns in Receivers' Actions

Considering that trust is the norm in the Trust Game—hence, the name—the role of the receiver is largely to reward and punish trusting behavior. The most pronounced patterns in the data are related to this reciprocal role. In the both the ID-only and S-ID versions of the experiment there was significant identity-related variation in reciprocity.

Proportion of Maximum Sent						
	ID-only			S-ID		
	overall	to in-group	to out-group	overall	to in-group	to out-group
Overall	0.513	0.539	0.487	0.553	0.576	0.531
	(0.022)	(0.031)	(0.031)	(0.019)	(0.027)	(0.027)
High Stat Sender				0.609***	0.651	0.569
				(0.026)	(0.037)	(0.037)
Low Stat Sender				0.495	0.500	0.490
				(0.028)	(0.039)	(0.040)

1. Reported value is $\$Sent$ as a proportion of maximum (i.e., $\frac{\$Sent}{7}$)
2. Standard errors in parentheses
3. Significance comparison is with respect to low-status senders in S-ID, and overall ID-only senders.

Table 5: Trust Game Senders' Actions

First of all, in the S-ID version of the experiment in which status was unequal, there is evidence of the hypothesis that high status subjects hold others to higher standards. To demonstrate this, I investigated reciprocity directly by estimating receivers' average return ratios as a function of the amount initially sent to them by their co-players. I used return ratios as the dependent variable in the estimation to avoid much of the heteroskedasticity associated with estimating return *amounts* directly.

To pin down a reasonable functional form, I separately estimated return ratio functions for high status and low status receivers as quintic functions of the amount initially sent. Wald tests failed to reject the null hypothesis that these return ratio functions were in fact linear.[8] Given this, Table 6 presents linear estimates of S-ID subjects' return ratio functions. The estimates reveal that high status receivers acted in a more reciprocal manner: compared to low status receivers, high status receivers' return ratios were more than twice as responsive to the amount initially sent—in some of the estimations, high status receivers were more than three times as reciprocal by this measure. Furthermore, the patterns proved robust to including individual receiver fixed effects; to adjusting the variance estimates by clustering on individual receivers; and to accounting for censoring by using a

[8] This was true for Wald tests on all subsets of powers of amount received greater than one as well

Dependent Variable = Return Ratio				
	(1)	(2)	(3)	(4)
Constant	0.451^{***}	0.467^{***}	0.513^{**}	0.519^{***}
	(0.144)	(0.117)	(0.199)	(0.173)
$\$Sent$	0.056^{**}	0.041^{**}	0.043	0.040
	(0.027)	(0.019)	(0.035)	(0.030)
High Status Receiver	-0.355^{**}	-0.419^{**}	-0.447^{*}	-0.431^{*}
	(0.174)	(0.187)	(0.228)	(0.253)
High Stat Rec $\times \$Sent$	0.086^{***}	0.121^{***}	0.106^{**}	0.124^{***}
	(0.033)	(0.027)	(0.042)	(0.043)
High Stat Sender			-0.146	-0.122
			(0.288)	(0.205)
High Stat Sender $\times \$Sent$			0.030	0.007
			(0.054)	(0.042)
H.S. Sender × H.S. Rec'r			0.229	-0.006
			(0.362)	(0.294)
H.S. Sender × H.S. Rec'r $\times \$Sent$			-0.047	0.002
			(0.068)	(0.056)
Receiver Fixed Effects	no	yes	no	yes
N	355	355	355	355
R^2	0.107	0.102	0.108	0.100

1. Dependent variable, return ratio, can take values from 0 to 3.
2. Robust standard errors in Parentheses.
3. Estimates include only observation where $\$Sent > 0$.

Table 6: Trust Game receivers, S-ID

Tobit estimation procedure. The latter two estimates are presented in the appendix.[9]

A point worth noting is that, to receivers, the status level of senders did not matter. This makes it more difficult to explain the results by simple distributional equity concerns. That is, one obvious explanation for the results is that the status manipulation simply introduced an unequal wealth distribution, since the low status group "earned less" by not receiving refreshments. However, receivers apparently ignored whether they were returning money to a "poorer" subject.

That high status subjects are more rewarding of trust and more punish-

[9] Additionally, throwing in session fixed effects as an added specification test did not change the patterns qualitatively.

ing of lack of trust is clearer in Figure 1. Also clearer is one economically relevant consequence of the difference in reciprocity patterns. From a purely pecuniary standpoint, trusting a low status subject was a losing proposition: low status subjects' average return ratios were below the break-even value of 1 no matter how much was sent. On the other hand, it was possible to earn positive returns on one's "investment" by trusting high status receivers. This pattern has potentially important dynamic effects, as, over time, subjects could learn to trust high status individuals and distrust low status individuals. Since I tried to eliminate dynamic effects in this experiment, however, I cannot presently test this intuition.

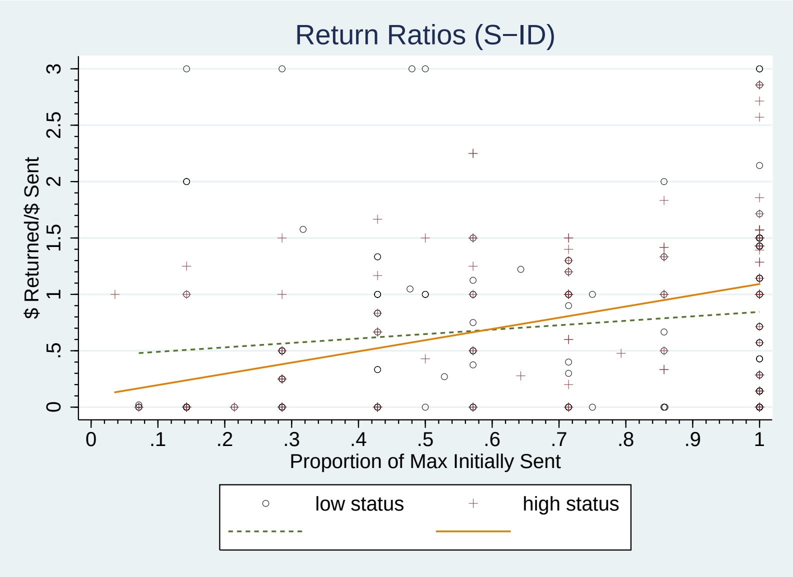

Figure 1: Trust Game Receivers, S-ID

Turning next to the ID-only Trust Game receivers, I estimated return ratio functions separately for subjects involved in same-color pairings vs. different-color pairings. Wald tests suggested the average return ratio function associated with same-color pairings was linear, while the function associated with different-color pairings was quadratic in money initially sent. Estimating return ratios using these functional forms—the estimates can be found in the appendix—revealed a significant difference between how in-group members and out-group members were treated: ID-only receivers held members of their *own* social identity to a higher standard. As is evident in Figure 2, when an in-group member sent a low amount they were punished more harshly than if they were not a member of the receiver's social identity; and when an in-group member sent a high amount, they were rewarded more generously than if they were not a member of the receiver's social identity.

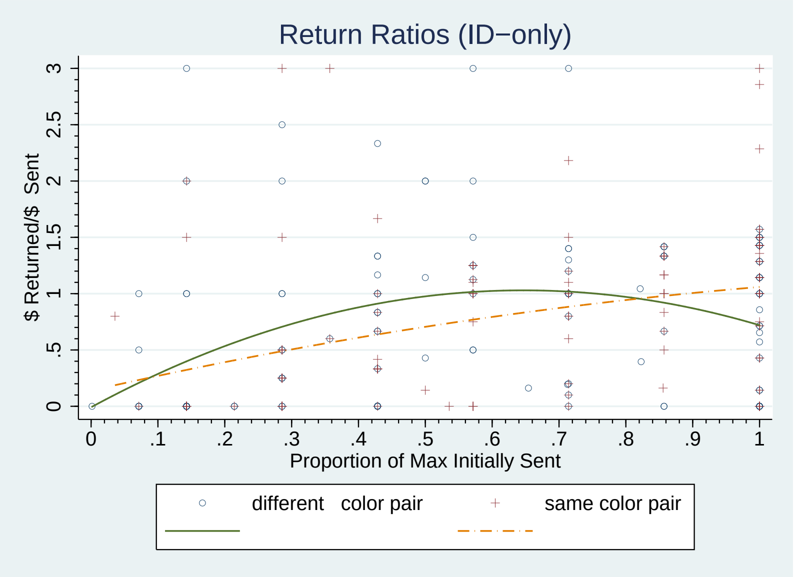

Figure 2: Return Ratios, ID-only

Comparing reciprocity patterns across the experimental conditions yields insights into the working of status. High status S-ID subjects held *all* subjects to a higher standard in the same way that ID-only subjects held members of their *own* identity to higher standards. At the same time, low status S-ID receivers were more reluctant to punish and reward trust in the same way that ID-only receivers were reluctant in punishing and rewarding members of the *other* identity.

In fact, these comparisons can be made more precise. Statistically, high status receivers in the S-ID version are indistinguishable from receivers involved in same-color pairings in the ID-only version: comparing the estimated return ratio functions for these two subsets of the data yielded no significant differences. That is to say, a Wald test failed to reject the null hypothesis of identical estimated return ratio functions ($p = 0.6460$). In the same respect, low status S-ID receivers' behavior was statistically indistinguishable from ID-only subjects engaged in *different-color* pairings ($p = 0.7488$). Furthermore, it is *not* the case that the data are just too noisy to distinguish among return ratio functions in any of these subsets:

high status receivers' estimated return ratio functions *were* significantly different from the the return ratio functions of ID-only subjects engaged in different-color pairings. A Wald test rejected the null hypothesis that return ratio functions were identical in these two cases ($p = 0.0259$).[10]

Putting these comparisons together, it seems as though status worked as follows: high status emboldened subjects to impose their values on *everyone* in the same way that subjects in the ID-only version were emboldened in their interactions with in-group members. Low status, on the other hand, makes subjects a bit more reserved in their judgments. Low status S-ID subjects withheld both punishment and reward in the same way ID-only receivers were less judgmental when dealing with members of the "other" group. In short, high status makes individuals more confident that their way is the right way, while low status makes individuals more humble.

4.2 The Truth Game

Next, I'll examine subjects' behavior in the Truth Game data in isolation. Afterwards, I'll draw connections between the patterns present in both games, and explain them in a social identity framework.

4.2.1 Summary Statistics

On the salesman side, the distribution of messages was surprisingly similar across versions—as Table 7 shows. Also, there was about a five percentage point increase in honesty in the S-ID version for each quality level: S-ID sellers are more likely to send the message "reliable" when the car is actually reliable; and more likely to send the message "lemon" when the car is actually a lemon.

For their part, buyers apparently conditioned their decisions on what sellers had to say, even though sellers' messages must reveal nothing about car quality in any purely self-interested equilibrium. In both versions of the experiment, using a Kolmogorov-Smirnov test, we can reject the idea that the distributions of buyers' actions were the same for both possible messages ($p < .01$). Buyers' actions are summarized in Table 8

[10] Reported Wald test results are when all return ratios were estimated as linear. The Wald tests were also conducted after all return ratios were estimated as quadratic in $\$Sent$. The significance patterns did not change.

Proportion of Messages Sent				
	ID-only		S-ID	
	“reliable”	“lemon”	“reliable”	“lemon”
Car Quality				
Overall	0.790	0.210	0.796	0.204
Reliable	0.896	0.104	0.942	0.058
Lemon	0.705	0.295	0.657	0.343

Table 7: Truth Game Sellers’ Messages, overview

Buyers’ Propensity to Buy/Walk				
	ID-only		S-ID	
	Buy	Walk	Buy	Walk
Message				
Reliable	0.751	0.249	0.770	0.230
Lemon	0.333	0.666	0.255	0.745
Overall	0.663	0.337	0.665	0.335

Table 8: Truth Game Buyers, overview

4.2.2 Truth Game Sellers

There is evidence supporting the high status/high standards hypothesis among Truth Game Sellers. For sellers, the norm is clear: a “decent” person *should* tell the truth. Therefore, if high status sellers hold themselves to a higher standard, we would expect high status sellers to be more honest than low status sellers. In fact, this was true.

In the S-ID version of the experiment, high status sellers were significantly more honest than low status sellers. Furthermore, as in the Trust Game, this “extra” honesty extended across versions: high status S-ID sellers were also more honest than ID-only sellers (Table 9).

Further still, high status sellers in the S-ID version of the experiment were more honest irrespective of their private information. They were more honest when they knew (privately) the car was reliable, and more honest when they knew the car was a lemon. Considered separately, the former pattern even rose to the level of statistical significance (Table 10).

One aspect of Table 10 is puzzling at first glance: why would anyone lie when the car is actually reliable?[11] After all, sellers’ and buyers’ preferences

[11] While one obvious explanation for this would be inexperience with the game, the fact that this pattern persists in all rounds suggests something else was at play.

Sellers' Propensity to Tell the Truth				
	ID-only	S-ID		
		Seller Status		
		Overall	High	Low
Buyer Status				
Overall	0.563	0.635**	0.683***	0.590
	(0.029)	(0.022)	(0.031)	(0.032)
High		0.644	0.699	0.603
		(0.031)	(0.045)	(0.042)
Low		0.624	0.669	0.573
		(0.033)	(0.043)	(0.049)

1. Standard Errors in parentheses
2. Statistical significance is with respect to ID-only
3. High status vs low status sellers overall: $p = 0.038$

Table 9: Truth Game Sellers' Honesty

Honesty Propensity, by Private Information		
	Car Quality	
	Reliable	Lemon
Seller Status		
High	0.991***	0.361
	(0.009)	(0.046)
Low	0.892	0.328
	(0.030)	(0.042)
ID-only	0.896	0.295
	(0.027)	(0.036)

1. Standard errors in parentheses
2. Statistical significance comparison is within-column

Table 10: Truth Game Sellers' Honesty

are aligned in this instance. One would think that "holding oneself to a higher standard" should imply more honesty in the face of bad news. And while this pattern *is* present, it would be reassuring if this bit of extra honesty were statistically significant.

The answer lies in the equilibria of the Truth Game. There is, in fact, an equilibrium where lying about good news is rational behavior: the incredulous equilibrium. The gist of this equilibrium is that if sellers think buyers will think sellers are always lying, then self-interested sellers will lie about good news and tell the truth about bad news. Specifically, in this equilibrium sellers with good news lie; sellers with bad news tell the truth; the only message that buyers ever hear, therefore, is bad news so buyers learn nothing from sellers' messages.

There is some evidence in the data suggesting incredulous equilibria play a role in the observed propensity to lie about good news. To see this, label each sender who lied at least once when they knew (privately) they had a good car for sale a "cynical seller." Among the observations in which sellers had private information that the car was a lemon, truth-telling was significantly *more* frequent if the observation involved a cynical seller ($\chi^2(1) = 11.3314, p = 0.001$). And among the observations in which the seller had private information that the car was reliable, truth-telling was much *less* likely if the observation involved a cynical seller: 100% by definition among non-cynical sellers vs. 60.9% among cynical sellers. This is exactly the pattern we'd expect if some of these subjects were engaged in the incredulous equilibrium: honesty after bad news and dishonesty after good news.

Furthermore, there is an apparent status bias in seller cynicism. In the S-ID data, of the 10 senders and 65 observations involved, only 1 sender—involved in 5 observations—was high status. Apparently, low status sellers expected to be disbelieved and developed a crude strategem to get their way.

Now, excluding all observations involving cynical sellers is clearly too blunt an instrument—as, for instance, this eliminates all the variation in honesty after good news. However, it is interesting to note that even this blunt tool reveals evidence in favor of the "high status/high standards" hypothesis: excluding cynical sellers, high status subjects are significantly more honest about bad news than both low status sellers and ID-only sellers (Table 11). Such a a pattern is more along the lines of what most people would consider "holding yourself to a higher standard."

Honesty about "Bad News"

	Seller Status	
ID-only	High	Low
0.266	0.364*	0.257
(0.037)	(0.047)	(0.044)

1. Standard errors in parentheses

Table 11: Honesty, Excluding "Cynical Sellers"

4.2.3 Truth Game Buyers

It's not clear what a norm for Truth Game buyers would be. And, since buyers in the Truth Game have no real punishment opportunity, it's not clear what would constitute evidence of holding others to a higher standard. Therefore, one wouldn't expect buyers' actions to change much across versions—and they don't (Table 8). But the Truth Game buyers still provide valuable insights. Specifically, I'll use the Truth Game buyers' data to highlight one consequence of introducing status (S-ID) relative to a "different-but-equal" (ID-only) environment: buyers unambiguously benefitted from the introduction of unequal status.

First, consider the purely self-interested equilibria of the Truth Game. In all self-interested equilibria of the Truth Game, buyers learn nothing from sellers' messages. Hence, we would expect buyers to choose the correct action—buy reliable cars and walk away from lemons—no more frequently than random chance allows (here, 50 percent of the time). This is exactly what happened among ID-only buyers—there was no evidence that buyers learned anything from sellers' messages. However, S-ID buyers *were* significantly more likely to choose the correct action than random chance would suggest (Table 12). Thus, the introduction of unequal status improved the functioning of this simple used-car market.[12]

How did this improvement happen? Notice in Table 12 that, overall, S-ID buyers did not fare significantly worse against low status sellers. Coupling the overall credulity evident among buyers in Table 8 with low status sellers' pronounced dishonesty, this seems odd. One obvious answer is that sellers somehow anticipated the honesty differential between high status and low

[12] As a specification check, I tested whether it was senders' messages that receivers conditioned their actions on, or if somehow senders' private information was revealed irrespective of messages—indicating a flaw in the experimental design. A multinomial logit estimation revealed no significant effects of the actual private information once messages were controlled for. The estimation appears in the appendix.

Buyers' Propensity to Choose "Correct" Action				
	ID-only	S-ID		
			Buyer Status	
Seller Status		Overall	High	Low
Overall	0.490	0.600***	0.594	0.606
	(0.029)	(0.023)	(0.032)	(0.033)
High Stat Seller		0.624	0.641	0.610
		(0.033)	(0.048)	(0.045)
Low Stat Seller		0.577	0.559	0.602
		(0.032)	(0.043)	(0.048)

1. Standard errors in parentheses
2. Indicated significance is with respect to overall ID-only
3. All S-ID values except (low-seller+low-buyer) are significantly different from .5 at $\geq 95\%$ confidence level

Table 12: Buying Reliable Cars and Walking Away from Lemons

status sellers. This seems to be exactly what happened: in the S-ID version of the experiment, overall, buyers were significantly more likely to believe messages from high status sellers than from low status sellers. This pattern in belief held for both high status and low status buyers, considered separately (Table 13).

Buyers' Propensity to Believe Messages			
	S-ID		
	Buyer Status		
Seller Status	Overall	High	Low
High	0.814**	0.825	0.805
	(0.026)	(0.037)	(0.037)
Low	0.720	0.721	0.718
	(0.029)	(0.039)	(0.045)

1. Standard errors in parentheses
2. Statistical significance is within-column

Table 13: Truth Game Buyers' Credulity (S-ID)

At the same time, ID-only buyers correctly gleaned the *lack* of variation in honesty. ID-only buyers were (correctly) no more likely to believe in-group members than out-group members (Table 14).

Considered as a whole, patterns in Truth Game buyers' actions provide

Sellers' Honesty and Buyers' Credulity		
	ID-only	
	Honesty Propensity (Sellers)	Belief Propensity (Buyers)
Type of Pairing		
Same Colors	0.553	0.713
	(0.041)	(0.037)
Different Colors	0.573	0.753
	(0.041)	(0.035)

1. Standard errors in parentheses

Table 14: Honesty and Credulity (ID-only)

evidence about the value of unequal status to the functioning of markets with asymmetric information and no verification opportunities. Buyers correctly anticipated the patterns in seller honesty implied by the introduction of status, and were able to use this inference to their advantage; thereby increasing the likelihood of mutually beneficial trades regardless of who the seller was.

5 A Social Identity Explanation

The patterns in the data can be succinctly explained with a simple identity model. For example, adding a concern for honesty to agents' utility functions turns the Truth Game into a costly signalling game—agents lose utility whenever they lie. If high status sellers lose more utility when they lie than low status sellers—i.e., they hold themselves to a higher standard—then, in equilibrium, we would expect more honesty from high status agents. If everyone knows this, we would also expect buyers to be more likely to believe high status sellers than low status sellers.

What remains to be seen is that social identity provides a *reasonable* explanation for the patterns. That is, even though the patterns are consistent with social identity in general, this wouldn't be a very satisfying explanation if the utility function, and in particular, the ideals, necessary to explain the data were ludicrous on their face. To demonstrate that social identity provides a reasonable explanation for the results, I estimated subjects' ideals for a subset of the observations in the data. The subset I chose involved Trust Game receivers in the S-ID version of the experiment, as this is where I have the most data and where the strategic situation is the most straightforward:

Trust Game receivers face a Dictator Game situation, so their decision of how much to return is purely a choice between ideals and self-interest.

Recall that in a social identity framework agents put themselves and others into social categories. Social categories are identities. Identities prescribe ideals—specifically, how agents *should* act. Since agents care about living up to their ideals and ideals are enmeshed with categorizations, in this type of model the *same* person can behave as if they are maximizing a different utility function when social identities change, even though nothing else decision-relevant has changed.

Reflecting these motivations, I constructed a simple identity model for Trust Game receivers' utility. Denoting senders' actions by s, receivers' actions by r and money payoffs by x, agent $j's$ overall utility is given by the following function, where α_j is a parameter capturing how much agent j cares about her identity relative to pecuniary incentives:

$$U_j = x_j(r,s) - \alpha_j(r - r^{Ideal}_{c_j}(s))^2 \tag{2}$$

In Equation 2, agent j cares about both her money earnings—$x_j(r,s)$—and the distance between her actual action, r, and her *ideal* action, $r^{Ideal}_{c_j}(s)$. Agent $j's$ ideal, in turn, depends on s—to incorporate reciprocity—as well as her social category, c_j. Of course, agent $j's$ ideal could also depend on her co-player's social category, but since there was no evidence of this in the S-ID version of the experiment, this possibility is not presently modeled.

While this model of utility might look unfamiliar, it, in fact, can be thought of as a generalization of a widely-used form of social preferences: inequity aversion (Fehr and Schmidt, 1999). To see this, consider a non-linear variant of the standard Fehr and Schmidt formulation of inequity aversion, which is also a close cousin to the prototypical example used in Bolton and Ockenfels (2000):

$$U_j = x_j(r,s) - \alpha_j(x_j(r,s) - x_i(r,s))^2 \tag{3}$$

Thus, agent j's cares about both her money earnings, and how unequal the distribution of earnings is. In the specific Trust Game used in the experiment, receivers' money earnings are given by $3s - r$ and senders' earnings are given by $7 - s + r$. Plugging these facts into (3) and simplifying yields:

$$U_j = 3s - r - \alpha_j(4s - 2r - 7)^2 \tag{4}$$

From Equation 4, simple algebraic manipulation allows one to re-cast the model of inequity aversion given by Equation 3 in an identity-utility form:

$$U_j = x_j(s,r) - \tilde{\alpha}_j(r - r^{Ideal}(s))^2 \qquad (5)$$

In Equation 5, $r^{Ideal}(s) = 2s - \frac{7}{2}$ and $\tilde{\alpha} = 4\alpha$.[13] Thus, inequity aversion can be thought of as an identity model with the additional assumption that ideals are constant across social identities, coupled with a specific prediction about what receivers' ideals are. Because of this connection, I will use as my null hypothesis the idea that the data in the current experiment are well-explained by inequity aversion.

Proceeding with the estimation of subjects' ideals, recall that in the current context, receivers' money payoffs in the Trust Game were $3s - r$. After plugging this fact into the identity model of Equation 2, first-order conditions imply that receivers' optimal (interior) money-return rule was given by:

$$r_j^*(s) = r_{c_j}^{Ideal}(s) - \frac{1}{2\alpha_j} \qquad (6)$$

Thus, estimating a subjects' average optimal return rule, $r^*(s)$, is equivalent—up to an unknown constant—to estimating subjects' ideals. To simplify matters, I assumed that ideals were linear in s and that there was precisely one ideal for each status level—$r_L^{Ideal}(s)$ for low status receivers, and $r_H^{Ideal}(s)$ for high status receivers. All individual-level heterogeneity, then, comes from the α_j's. To further simplify matters, I assumed that all α_j's were drawn from the same underlying distribution—as would be expected if this parameter captures a stable, individual trait. Call this random variable α.

The most straightforward way to estimate r^* was to use Equation 6 directly as a regression function, with $\frac{1}{2\alpha_j}$ serving as an error term. Since r^* was possibly censored—whenever r^* falls below zero, I observed zero[14]—and it's not clear what distribution the error term should have, I used a semi-parametric estimator that accounts for censoring and is robust to a wide range of error distributions: Censored Least Absolute Deviations (CLAD) (Powell, 1984).[15] It was sufficient, for instance, to assume that $\frac{1}{2\alpha}$ is a

[13] It looks rather out of place, but the 4 in $\tilde{\alpha}$ is an artifact of factoring -2 out of $(4s - 2r - 7)$ to get the expression in the parentheses, $(4s - sr - 7)$, into an $r - r^{Ideal}(s)$ format. This is also where the $\frac{7}{2}$ term in r^{Ideal} comes from.

[14] Censoring from above is also possible, but not quite as worrying as there are very few observations in the data where the maximum possible amount was returned.

[15] The main error-term assumption required for CLAD to be consistent is that errors have median zero, which is quite a bit less restrictive than the standard assumption of normality and homoskedasticity. The tradeoff is that an assumption must be made about

Dependent Variable = Dollars Returned				
	CLAD (1)	CLAD (2)	Tobit (1)	Tobit (2)
Constant	−0.60	-0.67	-2.18^{*}	-3.53^{**}
	(0.688)	(0.789)	(1.145)	(1.378)
$\$Sent$	0.80^{***}	0.67^{**}	0.96^{***}	1.219^{***}
	(0.297)	(0.322)	(0.225)	(0.253)
High Status Receiver	-6.90^{***}	-3.33	−2.76	−2.62
	(2.369)	(2.805)	(1.687)	(2.063)
H. S. Rec'r $\times \$Sent$	1.70^{***}	1.33^{**}	0.70^{**}	0.72^{*}
	(0.541)	(0.609)	(0.316)	(0.376)
High Stat Sender		-0.33		-0.13
		(1.726)		(2.100)
H.S. Sender $\times \$Sent$		0.33		0.07
		(0.612)		(0.379)
H.S. Rec'r × H. S. Sender		1.00		0.52
		(3.998)		(3.237)
H.S. Rec'r × H. S. Sender $\times \$Sent$		-0.67		-0.194
		(0.789)		(0.568)
$\mathrm{N}_{initial}$	355	355	355	355
N_{final}	302	315	355	355
Pseudo R^2	0.187	0.216	0.090	0.068

1. Standard errors in Parentheses

2. Estimates include only observation where $\$Sent > 0$

Table 15: Optimal Return functions, S-ID

well-defined random variable with a unique (finite) median.

The CLAD estimates of the optimal return functions are presented in Table 15. The estimated return functions are significantly different across status levels—allowing us to rule out inequity aversion. In addition to CLAD, Table 15 includes Tobit estimations for comparison. However, the regularity conditions necessary for the consistency of maximum likelihood estimation are not obviously satisfied, so the Tobit estimates may be biased.

Since none of the controls related to senders' status are significant—neither individually nor jointly—I will focus on the simplest CLAD estimates from here on. Call the estimated optimal return rule among high status subjects $\widehat{r_H^*}$, and define $\widehat{r_L^*}$ analogously. Estimating receivers' ideals is a

the data—roughly speaking, there must be "enough" uncensored observations. As will be clear from the estimation, this is likely to be satisfied in the present case.

matter of simply shuffling terms from one side of Equation 6 to the other. Specifically:

$$\widehat{r_H^{Ideal}} = \widehat{r_H^*} + \frac{1}{2\overline{\alpha}} \tag{7}$$

$$\widehat{r_L^{Ideal}} = \widehat{r_L^*} + \frac{1}{2\overline{\alpha}} \tag{8}$$

Plugging the CLAD estimates, $\widehat{r_H^*}$ and $\widehat{r_L^*}$, into Equations 7 and 8 yields the estimated ideals:

$$\widehat{r_H^{Ideal}} = 2.5s - 7.5 + \frac{1}{2\overline{\alpha}}$$

$$\widehat{r_L^{Ideal}} = 0.8s - 0.6 + \frac{1}{2\overline{\alpha}}$$

Notice that high status subjects' ideals are not terribly different from what would be expected from inequity-averse agents. Further, recall that high status subjects act just like subjects in the ID-only version of the experiment engaged in in-group pairings. This suggests one reason for the frustrating lack of external validity characteristic of many previous experimental investigations of trust and reciprocity: merely bringing subjects into the lab succeeded in creating a shared social identity.

5.1 An Interpretation of the Estimated Ideals

One way to think of the receiver's role in the Trust Game is as an enforcer of normative behavior—punishing deviations from the norm and rewarding norm conformance. With this in mind, and denoting the sender's norm with s^{Ideal}, we can re-write receivers' estimated ideals in a particularly simple format (Equation 9). Here, γ captures a base-line level of generosity, while β measures concern for others living up to their ideals:

$$r^{Ideal}(s) = \beta(s - s^{Ideal}) + \gamma \tag{9}$$

To make this more concrete, suppose that in the Trust Game all senders *should* exhibit full trust—i.e., $s_H^{Ideal} = s_L^{Ideal} = 7$. Then the estimated ideals can be re-written in an especially tidy manner:

$$r_H^{Ideal}(s) = 2.5(s - s^{Ideal}) + (10 + \frac{1}{2\overline{\alpha}}) \tag{10}$$

$$r_L^{Ideal}(s) = 0.8(s - s^{Ideal}) + (5 + \frac{1}{2\overline{\alpha}}) \tag{11}$$

Here we see receivers' true colors. High status receivers care much more about others' norm conformance—punishing senders by more than two dollars for every dollar senders fall short of their ideal. But they are also much more generous when it is warranted. This last point can be seen by considering the ideal return amount when senders exactly conform to their norm. In this case, $(s - s^{Ideal}) = 0$, implying that high status receivers (ideally) reward senders much more lavishly than low status receivers.

5.2 Ruling out Wealth Effects and Mood Effects

One obvious alternative explanation for the patterns in the data is some type of wealth effect: it could be that the status manipulation in the S-ID version of the experiment simply made the "high status" subject feel wealthier. The exact effect of extra wealth would depend on the model used, but heuristically one might expect the "wealthier" subjects to be uniformly more generous—returning a fixed amount more than their poorer counterparts. Since this constant extra generosity would have greater impact on return *ratios* for low return amounts, this uniform extra generosity should imply *flatter* return ratio functions for high status/wealthier subjects. This was exactly the *opposite* of the observed patterns. Furthermore, it would be difficult to explain the connection between the ID-only patterns and the S-ID patterns with wealth effects alone, since it would be difficult to make the case for a systematic wealth difference in the ID-only version of the experiment. Finally, while the increased generosity of high status subjects when sent a large initial amount is consistent with wealth effects, the relative *decrease* in generosity observed when the "wealthier" subjects were sent low amounts seems contrary in spirit to the idea that high status subjects simply feel wealthier.

Another possible explanation for the patterns in the data is a mood effect. One might think that the manipulation I used to reinforce status simply put half the subjects—the high status subjects—in a better mood. And, there is some experimental research in economics suggesting that mood might have an effect on "other-regarding" behavior. I again, however, have a few reasons for believing mood effects do not explain the results. Firstly, I will reiterate that social identity had a significant impact even in the ID-only version of the experiment where it's less clear that moods varied. Secondly, in a companion paper, I demonstrate effects on reciprocity similar to those in the S-ID version, using a manipulation intended to make identity more salient, but also unlikely to have pronounced mood effects: having subjects write briefly (10 minutes) about the values important to themselves (version

1) or to someone else (version 2) (Butler, in preparation).

Thirdly, my own work aside, while clean results are hard to come by with respect to mood's impact on behavior, there is one experiment in this vein that provides compelling evidence that mood does not explain the current results. In Kirchsteiger, Rigotti and Rustichini (2001), the authors induced two different moods—good or bad—by having subjects watch either a funny or a sad movie, respectively. Subjects then played a standard gift-exchange game. A gift exchange game is similar to the Trust Game. There are two players. One player moves first and chooses how much of a fixed sum of money to transfer to his or her co-player—player 2. Player 2 observes the amount transferred and decides how much "effort" to exert. Effort decreases player 2's earnings but increases player 1's earnings. Reciprocity is measured as the responsiveness of the effort decision to the transfer decision. Generosity is measured by the amount of effort exerted when the initial transfer is zero. The authors found that subjects in a good mood are *less* reciprocal and *more* generous than players in a bad mood. That is to say, players in a *bad* mood mirror high status subjects in the present experiment; and players in a *good* mood mirror low status subjects.[16] Thus, mood differences are likely to have worked *against* the reciprocity patterns in the current paper, rather than providing an alternative explanation for them.

6 Conclusion

On the most basic level, I have shown that subjects' ideals vary with social identity, rather than being stable *individual* traits. Beyond this, the results provide evidence for specific effects of simple social identities that are reasonable and ring true: we hold our own people to higher standards; and high status emboldens us to hold everyone—including ourselves—to higher standards.

The social identity effects above tie together results across unrelated experiments in a tidy manner—trustworthiness patterns in Glaeser, Laibson, Scheinkman and Soutter (2000) and Fershtman and Gneezy (2001), as well as price patterns in Ball, Eckel, Grossman and Zame (2001). At the same time, the results provide one plausible factor—the unintended creation of a shared social identity—in the frustrating lack of external validity associated with many social preferences experiments in economics.

Additionally, the current experiment raises many questions for future research. First of all, the social identities induced in this experiment were

[16] They also mirror, by the way, the hypothetically "wealthier" subjects outlined above.

assigned rather than chosen, and status was assigned rather than earned. This opens the question of whether the effects of social identity and status generalize to earned status and/or chosen social identities.

Secondly, the current experiment was designed to shut down dynamic effects. Opening up the investigation of social identity to such dynamic effects raises potentially interesting questions. Most directly, patterns in the Trust Game suggest that subjects may learn, over time, to trust only high status individuals, as completely trusting high status individuals was the only way to earn a positive expected return-on-investment.

Another class of questions raised by the current experiment relates to multiple social identities. In this experiment I attempted to induce unique identities. How individuals prioritize multiple social identities, particularly when these identities have conflicting ideals, is an obvious extension that would provide a closer analogy to the real world.

Finally, the current experiment dictated the associational patterns: with whom subjects interacted was (randomly) assigned. However, one might suspect that, for instance, the patterns in punishing norm-deviance might have a significant impact on with whom subjects *choose* to interact when these associational decisions are voluntary. To the extent that social networks are economically important, such associational biases may have significant economic consequences, and are thus worthy of investigating.

References

[1] George A. Akerlof. Labor contracts as partial gift exchange. *Quarterly Journal of Economics*, 97(4), 1982.

[2] George A. Akerlof and Rachel E. Kranton. Economics and identity. *Quarterly Journal of Economics*, CVX(3), August 2000.

[3] Kenneth Arrow. Gifts and exchanges. *Philosophy and Public Affairs*, 1:343–362, 1972.

[4] Sheryl Ball, Catherine Eckel, Phillip Grossman, and William Zame. Status in markets. *Quarterly Journal of Economics*, 116(1):161–188, February 2001.

[5] Roland Benabou and Jean Tirole. Incentives and prosocial behavior. 2004.

[6] Joyce Berg, John Dickhaut, and Kevin McCabe. Trust, reciprocity and social history. *Games and Economic Behavior*, 10:122–142, 1995.

[7] Truman F. Bewley. *Why Wages Don't Fall During A Recession.* Harvard University Press, Cambridge, Mass, 1999.

[8] Gary E. Bolton and Axel Ockenfels. A theory of equity, reciprocity and competition. *American Economic Review*, 100:166–193, 2000.

[9] Roger Brown. *Social Psychology.* The Free Press, New York, 1965.

[10] Rupert Brown. Social identity theory: past achievements, current problems and future challenges. *European Journal of Social Psychology*, 30:745–778, 2000.

[11] Jeffrey V. Butler. Identity salience and reciprocity. in progress, 2007.

[12] Colin F. Camerer. *Behavioral Game Theory: Experiments in Strategic Interaction.* Princeton University Press, Princeton, New Jersey, 2003.

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

[14] Ernst Fehr and Klaus M. Schmidt. A theory of fairness, competition and co-operation. *Quarterly Journal of Economics*, 114(3):817–868, August 1999.

[15] Chaim Fershtman and Uri Gneezy. Discrimination in a segmented society: An experimental approach. *Quarterly Journal of Economics*, 116(1):351–357, February 2001.

[16] Urs Fischbacher. z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2):171–178, 2007.

[17] Luca Rigotti Georg Kirchsteiger and Aldo Rustichini. Your morals are your moods. *Journal of Economic Behavior and Organization*, 59(2):155–172.

[18] Edward Glaeser, David I. Laibson, Jose A. Scheinkman, and Christine L. Soutter. Measuring trust. *Quarterly Journal of Economics*, 115(3):811–846, 2000.

[19] S. Alexander Haslam. *Psychology in Organizations: The social identity approach.* Sage Publications, Ltd., Thousand Oaks, CA, 2001.

[20] George Caspar Homans. *The Human Group.* Harcourt, Brace, New York, 1950.

[21] S. Knack and P. Keefer. Does social capital have an economy payoff: A cross-country investigation. *Quartely Journal of Economics*, pages 1251–1288, November 1997.

[22] David Levine. Modeling altruism and spitefulness in experiments. *Review of Economic Dynamics*, (1):593–622, 1998.

[23] Brian Mullen, Rupert Brown, and Colleen Smith. Ingroup bias as a function of salience, relevance and status: an integration. *European Journal of Social Psychology*, 22:103–122, 1992.

[24] J L Powell. Least absolute deviations estimation for the censored regression model. *Journal of Econometrics*, 25:303–325, 1984.

[25] Fritz J. Roethlisberger and Willam J. Dickinson. *Management and the Worker: An Account of a Research Program Conducted by the Western Electric Company, Hawthorne Works, Chicago.* Harvard University Press, Cambridge, MA, 1939.

[26] Amartya Sen. *Identity and Violence: the illusion of destiny.* W. W. Norton and Company, New York, New York, 2007.

[27] M. Sherif, O.J. Harvey, B.J. White, W.R. Hood, and C.W. Sherif. *Intergroup conflict and cooperation: The Robbers Cave experiment.* University Book Exchange, Norman, Oklahoma, 1954.

[28] Todd G. Shields and Robert K. Goidel. Participation rates, socioeconomic class biases, and congressional elections: A crossvalidation. *American Journal of Political Science*, 41(2):683–691, April 1997.

[29] H. Tajfel, C. Flament, M.G. Billig, and R.F. Bundy. Social categorization and intergroup behavior. *European Journal of Social Psychology*, 1:149–177, 1971.

[30] Henri Tajfel and Michael Billig. Social categorization and similarity in intergroup behavior. *European Journal of Social Psychology*, 3(1):27–52.

[31] J. Turner and R. Brown. *Differentiation Between Social Groups*, chapter Social status, cognitive alternatives and intergroup relations. Academic Press, London, 1978.

A Figures

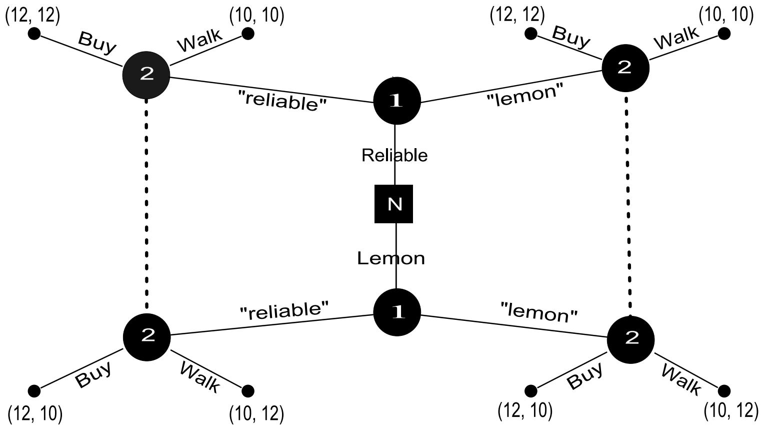

Figure 3: Truth Game, game tree

B Tables

Dependent Variable = Return Ratio				
Variable	(1)	(2)	(3)	(4)
Constant	0.451^{**}	0.467^{***}	0.469^{**}	0.520^{***}
	(0.144)	(0.134)	(0.224)	(0.162)
$\$Sent$	0.056^{**}	0.041^{*}	0.052^{*}	0.039
	(0.023)	(0.022)	(0.028)	(0.028)
High Status Receiver	-0.355^{*}	-0.419^{*}	-0.357	-0.433^{*}
	(0.213)	(0.222)	(0.218)	(0.219)
High Stat Rec $\times\$Sent$	0.086^{***}	0.121^{***}	0.086^{***}	0.125^{***}
	(0.031)	(0.032)	(0.031)	(0.032)
High Stat Sender			-0.043	-0.125
			(0.162)	(0.143)
High Stat Sender $\times\$Sent$			0.008	0.008
			(0.033)	(0.028)
Receiver Fixed Effects	no	yes	no	yes
N	355	355	355	355
R^2	0.107	0.102	0.108	0.100

1. Dependent variable, return ratio, can take values from 0 to 3
2. Robust standard errors in Parentheses
3. Errors adjusted to include intra-receiver variation (clustering)
4. Estimates include only observation where $\$Sent > 0$

Table 16: Trust Game receivers, clustered, S-ID

Dependent Variable = Return Ratio				
Variable	(1)	(2)	(3)	(4)
Constant	0.052	0.081	0.021	0.010
	(0.189)	(0.174)	(0.226)	(0.194)
$\$Sent$	0.098^{***}	0.074^{***}	0.103^{**}	0.079^{**}
	(0.034)	(0.027)	(0.041)	(0.032)
High Status Receiver	-0.609^{**}	-0.661^{***}	-0.606^{**}	-0.670^{***}
	(0.286)	(0.245)	(0.287)	(0.244)
High Stat Rec $\times\$Sent$	0.131^{**}	0.156^{***}	0.131^{**}	0.159^{***}
	(0.051)	(0.039)	(0.051)	(0.039)
High Stat Sender			0.074	-0.037
			(0.289)	(0.220)
High Stat Sender $\times\$Sent$			-0.013	0.008
			(0.051)	(0.040)
Receiver Random Effects	no	yes	no	yes
N	355	355	355	355

1. Dependent variable, return ratio, can take values from 0 to 3

2. Standard errors in Parentheses

3. Estimates include only observation where $\$Sent > 0$

Table 17: Trust Game receivers, Tobit, S-ID

Dependent Variable = Return Ratio		
Variable	(1)	(2)
Constant	-0.009	-0.009
	(0.199)	0.211
\$ Sent	0.459***	0.459***
	(0.121)	(0.126)
$(\$Sent)^2$	$-.051^{***}$	0.051^{***}
	(0.015)	(0.016)
Same-Color Pairing	0.149	0.149
	(0.306)	(0.258)
Same-Color Pairing $\times\$Sent$	-0.267	-0.267^{*}
	(0.174)	(0.152)
Same-Color Pairing $\times(\$Sent)^2$	0.042**	0.042**
	(0.021)	(0.019)
Receiver Clustering	no	yes
N	249	249
R^2	0.126	0.126

1. Dependent variable, return ratio, can take values from 0 to 3

2. Robust standard errors in Parentheses

3. Estimates include only observation where $\$Sent > 0$

Table 18: Trust Game receivers, ID-only

Dependent Variable = Buyer's Action		
	Logit (1)	Logit (2)
Message	-1.83^{***}	-2.17^{***}
	(0.595)	(0.824)
Car Quality	-0.340	-0.466
	(0.250)	(0.354)
Quality × Message	-0.37	-0.248
	(0.675)	(0.890)
Constant	1.36***	—
	(0.171)	
Rec'r Fixed Effects	no	yes
N	355	355
Pseudo R^2	0.151	0.289

Table 19: Specification Test: Truth Game Buyers

C Instructions

The following instructions appeared on each subject's computer screen immediately before each game was started:

C.1 The Trust Game

In each round of the following game, you will be randomly paired with a co-player and will be randomly assigned one of two roles: **Sender** or **Receiver.**

Role of Sender:
The **Sender** will be given \$7.00 and may choose to send any amount of this money to the **Receiver**. The amount chosen by the **Sender** will be tripled and given to the **Receiver**.

Role of Receiver:
The **Receiver**, upon learning how much money is available to him or her, will be able to send any amount of that money back to the **Sender**.

Earnings per round:
The **Sender** earns 7.00 - (dollars sent to Receiver) + (dollars sent back by Receiver);

The **Receiver** earns 3∗(dollars sent by Sender) - (dollars sent back to Sender).

When you've read and understand the Instructions, click on the button to proceed to the game.

C.2 Truth Game

Now we begin a new game.

As before, in each round of the following game, you will be randomly matched with a co-player and randomly assigned one of two roles: **Sender** or **Receiver**.

Role of Sender: The **Sender** will learn the result of a coin flip performed by the computer. This coin flip has equal chances of resulting in "Heads" or "Tails." The **Sender** then sends a message about this coin flip to the

Receiver.

Role of Receiver: The **Receiver**, does not learn the result of the coin flip. The **Receiver's** only information about the result is the message sent by the **Sender**. Once the **Receiver** receives this message from the **Sender**, the **Receiver** chooses an action: *Left* or *Right.* The **Receiver's** action, together with the result of the coin flip, completely determines both the **Sender's** and the **Receiver's** earnings for the round.

Earnings per Period: If the result of the Coin Flip is Heads, then the action Left yields earnings of $12 for the Receiver and $12 for the Sender. The action Right yields earnings of $10 for the Receiver and $10 for the Sender.

If the result of the Coin Flip is Tails, then the earnings for the action Left are $10, $12 for Receiver, Sender. And the earnings for the action Right are $12,$10 for Receiver, Sender.

This is summarized in the following table, where: $\$a, \$b$ = $a earnings for Receiver, $\$b$ earnings for Sender.

Coin/Action	Left	Right
Heads	$12, $12	$10, $10
Tails	$10, $12	$12, $10

Click the button when you're ready to proceed to the game.