Trust, Truth, Status and Identity:

an experimental inquiry *

Jeffrey V. Butler

Einaudi Institute for Economics and Finance

jeff.butler@eief.it

This Version: November 2013

Abstract: To investigate how group-contingent non-pecuniary preferences are affected when one group occupies a position of higher status than another group, experimental participants were divided into two trivially distinct groups and then one of the groups was randomly assigned "high status." Control sessions were also conducted in which no status distinction was introduced. In all sessions participants subsequently played two games governed by distinct social norms: a trust game and a cheap talk game where lying was possible. In the control sessions norm compliance was higher in same-group interactions, consistent with previous research demonstrating that normative obligations are often parochial. In treatment sessions parochialism vanished and was replaced by "noblesse oblige:" members of high status groups exhibited more norm compliance in all of their interactions. Finally, in game roles not governed by an unambiguous social norm, identity had no direct impact on behavior. Considered together, the results suggest that the channel through which social identity directly impacts behavior is norm compliance and that the nature of this impact depends crucially on the relationship between involved groups.

*I owe a great debt of gratitude to George A. Akerlof, Stefano DellaVigna, Steven Marc Goldman, Luigi Guiso, Shachar Kariv, Ulrike Malmendier, John Morgan and David Reinstein for many helpful comments and suggestions. I also thank seminar participants at the University of California, Berkeley, Oregon State University, the Einaudi Institute for Economics and Finance (Italy) and the University of Trento (Italy). Finally, I am grateful to the National Science Foundation under Research Grant SES 04-17871 and UC Berkeley's X-lab for generous financial support.

1 Introduction

The potential for mutually beneficial exchange often arises in contexts where contracts are necessarily incomplete. Small stakes or high legal costs may mitigate against writing contracts at all; non-verifiable private information may make enforcing complete contingent contracts impossible. In such situations, realizing potential gains from trade depends on the non-pecuniary moral, ethical and normative considerations that regulate interpersonal interactions. Consequently, over the last two decades understanding the determinants of non-pecuniary preferences has come within the purview of economics. More recently, economists have become interested in the explanatory power of group affiliation in shaping ethical preferences.

Building on Social Identity Theory (Tajfel, et al., 1971), Akerlof and Kranton (2000) provide motivating examples and a theoretical model for incorporating group-contingent ethical preferences into the economic discourse. Several subsequent studies provide initial experimental evidence of the potential for social identity to enrich our economic models. In the bulk of these studies, participants are divided into trivially distinct groups—minimal or near-minimal—and then make decisions in games featuring a prominent role for non-pecuniary preferences (e.g., variants of the dictator game). Confirming and extending existing results in social psychology, a common finding across a variety of different games in this "equal-but-different-groups" context is in-group bias.[1] That is to say, participants tend to treat members of their own group better than other-group members, being, for example, more willing to share their own money or cooperate with in-group members than with out-group members (Eckel and Grossman, 2005; Charness, Rigotti and Rustichini, 2007; Chen and Li, 2009; Chen and Chen, 2011).

Having established the robustness of in-group bias across strategic contexts when groups are equal-but-different, the question naturally arises about whether in-group bias continues to be a factor when groups are unequal. After all, group-level status differences are a central and salient feature of many of the contexts economists care about: firms and organizations are comprised of both equal-status groups (workers in different plants) and unequal-status groups (workers vs. manage-

[1]However, there are a few important counter-examples and negative findings. In terms of counter-examples, McLeish and Oxoby (2008) find that individuals are *more likely* to levy costly punishment on in-group members for bad behavior than on out-group members, which seems inconsistent with treating in-group members better but is consistent with an evolutionary group-selection model for the existence of social norms (Bernhard, et al., 2006). In terms of negative results, several studies find no effect of equal-but-different groups at all: Güth, Levatti and Ploner (2008), as well many of the treatments in Charness, Rigotti and Rustichini (2007) and Eckel and Grossman (2005).

ment); public goods provision often requires cooperation among unequal groups (e.g., homeowners and renters); and charitable giving often takes place between social groups of unequal status. Consequently, understanding whether and how group-level status inequality affects group-contingent preferences is an important undertaking.

This study takes a first step towards understanding the role of unequal status in determining group-contingent preferences. This paper presents the results of a laboratory experiment consisting of treatment and control sessions. Both treatment and control featured an identity-inducement phase, followed by a game-playing phase. Only the identity-inducement phase varied from treatment to control. Specifically, in treatment sessions—S-ID—participants were publicly randomized into two trivially distinct groups and one of the two groups was then randomly assigned the label "high status." Subsequently, each group completed status-appropriate tasks. In the control sessions—ID-only—participants were also publicly randomized into two trivially distinct groups in a manner identical to that in the treatment sessions, but the identity inducement phase stopped there. That is to say, in control sessions no status distinction was introduced, nor were subsequent status-appropriate tasks implemented.

In both treatment and control sessions, after the identity-inducement phase was complete participants played two games: first a trust game (Berg, Dickhaut and McCabe, 1995) and then a two-player cheap talk game where lying is well defined and potentially profitable. For ease of exposition, I refer to this latter game as the "Truth Game." Details of both games are provided in the Experimental Design section, below.

These two games were chosen because each permits a prominent role for ethical, non-pecuniary, preferences in determining outcomes. At the same time, the ethical considerations which plausibly govern behavior differ across these two games and, within each game, by role. I refer to a broadly shared consensus about what constitutes good behavior—how people *ought* to act—as a social norm (*cf.*, Akerlof, 1980; Ostrom, 2002; Krupka and Weber, 2013; Kimbrough and Vostroknutov, 2013). Behavior in the trust game is widely believed to be governed by social norms of trust (sender) and reciprocity (receiver). In the truth game, honesty is a plausible social norm governing the behavior of sellers (see, e.g., Gneezy, 2005). Moreover, having the same participants play different games governed by different normative considerations will allow me to shed light on whether group distinctions or group-level status differences affect only distributional preferences or, rather, the

weight given to ethical considerations in general. The latter pattern is suggested by Homans (1950) and finds empirical support in Guiso, Sapienza and Zingales (2003),[2] but direct evidence is scant since most studies on group-contingent preferences involve only a single game or multiple games governed by similar ethical concerns (e.g., inequality aversion in money division tasks). Because the set of games is constant across treatment and control sessions, comparing behavior in the ID-only treatment to behavior in the the S-ID treatments allows me to isolate the effect that group-level status differences *per se* have on group-contingent preferences.

My results suggest that in-group bias is a manifestation of an enhanced concern for normative behavior in interactions involving in-group members when groups are trivially distinct. When groups differ in terms of status, however, in-group bias is replaced by a form of paternalism or *noblesse oblige*: high status group members behave more in accordance with relevant social norms in all of their interactions regardless of their co-player's group affiliation.

One way to make sense of these patterns is to recall that a central tenet of Social Identity Theory is that individuals adopt group/social identities in order to achieve "positive distinctiveness." When groups are trivially distinct, competition for group distinction may lead individuals to follow strategies that are individually costly but that improve the relative image or position of one's group in order to create the possibility for a group label to convey positive distinction. This can be accomplished directly through classical in-group bias. It can also be accomplished indirectly, however, by providing stronger incentives for in-group others to behave in a way that increases group success—e.g., by punishing in-group members more severely for violating welfare-enhancing norms. This "indirect" in-group bias is consistent with, and can give rise to, "parochial altruism" (Bernhard, et al., 2006) or its more extreme form, "amoral familism" (Banfield, 1958).[3]

On the other hand, when a status hierarchy already exists between groups and is not readily overturned, the distinctiveness motivation for in-group bias is weakened, allowing other facets of group-contingent preferences to dominate. A common expectation, rooted in widespread moral

[2]The World Values Survey asks how justifiable such morally wrong behaviors as cheating on taxes or unjustly claiming government benefits are. Respondents answer on a scale from 1 (never justifiable) to 10 (always justifiable). The authors find that respondents who view themselves as belonging to a higher social class tend to rate these behaviors as less frequently justifiable.

[3]This pattern is also consistent with the classical in-group bias results in the context of disinterested money division tasks, particularly if in-group bias *is* the norm (Harris, et al., 2010; G. Hertel and N. L. Kerr, 2001; Jetten, Spears and Manstead, 1997).

codes such as that presented by the New Testament[4] as well as popular culture[5] is that higher status groups *should* care more about normative behavior and, consequently, comply more closely to relevant norms. If individuals either directly internalize, and behave according to, this expectation, or care about it indirectly because of an aversion to disappointing others (Charness and Dufwenberg, 2006; Battigalli and Dufwenberg, 2007), what Homans termed *noblesse oblige* may obtain.

The remainder of this paper is organized as follows. First, I discuss closely related research. Next, the experimental design is presented in detail and a specific testable hypothesis is given. In the next section, results are presented—first for the trust game and then for the truth game. In Section 6, I argue that the results cannot be plausibly explained by either wealth or mood effects. In the subsequent section, I construct a simple model of social identity consistent with the data and estimate its parameters for a subset of the data where they are identifiable. In the last section I summarize the results and suggest future avenues of research. Experimental instructions appear in the Appendix.

2 Closely Related Literature

The most closely related research in experimental economics involves inducing two trivially different categories—"us" vs. "them"—in order to test whether this creates in-group bias.[6] Classical in-group bias often does arise (Chen and Li, 2009; Eckel and Grossman, 2005; Charness, Rigotti and Rustichini, 2007; Guala, Mittone and Ploner, 2013). An important exception to the robust finding of in-group bias in this context is provided by McLeish and Oxoby (2008), where the authors find that individuals are more likely to levy costly punishment on in-group members for bad behavior than on out-group members. Hargreaves-Heap and Zizzo (2009) show that individuals place a non-trivial monetary value on such group affiliations. Finally, several studies find no effect of such equal-but-different social identities at all (see, e.g., Güth, Levatti and Ploner, 2008).[7] The current paper differs from these studies by focusing not on trivially distinct groups directly, but rather on

[4]Two relevant examples: i) "From everyone who has been given much, much will be demanded; and from the one who has been entrusted with much, much more will be asked." Luke 12:48; ii) "Not many of you should become teachers, my fellow believers, because you know that we who teach will be judged more strictly." James 3:1

[5]Peter Parker's uncle, in Stan Lee's *Spiderman*, famously advises: "With great power comes great responsibility."

[6]The experimental literature in social psychology starts with Tajfel et al (1971). See Mullen et al (1992) or the references and discussion in Chen and Li (2009) for a nice overview of subsequent psychological research.

[7]Also many of the treatments in Charness, Rigotti and Rustichini (2007) and Eckel and Grossman (2005) yielded no significant evidence of in-group bias.

the difference in behavior that results when a group-level status difference is explicitly introduced between otherwise trivially distinct groups.

A second closely related strand of research takes in-group bias as given and asks what accounts for it: innate preferences or group-contingent variation in norm-concern. Harris, Hermann and Kontoleon (2010) show that in-group bias increases when there is opportunity for third-party punishment, suggesting that in-group bias is a social norm in the context of equal-but-different groups. This interpretation is strengthened by two other papers: Hertel and Kerr (2001) find that priming specific norms that imply in-group bias (e.g., group loyalty) increases in-group bias and the expectation of others' in-group bias; Jetten, Spears and Manstead (1997) find the effect of norm-priming on in-group bias is stronger for those who identify more strongly with their group. Differently from these papers, the current experiment varies the relevant norms, for the same individuals, across different games.

The current paper is also related to a handful of studies examining how uninformative, laboratory-induced, status differences affect behavior in situations where non-pecuniary incentives are important (Ball and Eckel, 1998; Kumru and Vesterlund, 2008; Tsutsui and Zizzo, *forthcoming*; Willer, 2009). The results in these studies are generally consistent with status increasing compliance with relevant norms. The main innovation of the current paper with respect to these studies is to implement multiple games involving roles governed by distinct social norms among the same participants in order to test if status magnifies the impact of social norms in general or, instead, whether only specific social norms such as altruism or generosity are affected. Moreover, in contrast to most of these studies by comparing behavior with and without unequal status groups my design allows me to isolate how the *introduction* of a status difference into a situation with pre-existing social groups affects behavior.

The current inquiry is also related to several papers in economics investigating the impact of real-world social identities, such as gender or ethnicity, on behavior (Fershtman and Gneezy, 2001; Hoff and Pandey, 2006; Benjamin, Choi and Strickland, 2010; Ben-Ner et al, 2009; Chen et al, 2010) and to a closely related body of work in social psychology using real-world group affiliations with a specific focus on groups of varying (real-world) status. Results using real-world social identities vary and are not easily summarized. In particular, in the social psychological literature the otherwise-robust finding of in-group bias in the context of equal-status groups is often reversed

when groups differ in terms of status—producing out-group bias. Why this happens is not well understood and was an early dilemma for Social Identity Theory that continues to be controversial. For an overview of the social psychological research in this vein, see, e.g., Brown (2000) or Jost, Banaji and Nosek (2004). The current paper differs from this line of inquiry by experimentally inducing social identities and status in an attempt to avoid some of the confounds (e.g., stereotypes and cultural differences) associated with dividing participants by their actual characteristics.

Finally, while neither the existing social psychological nor economics literature provide clear predictions about the effects of group-level status differences on group-contingent behavior, a classic study by (eminent sociologist) Homans (1950) does. Observing the interactions between two groups of workers whose jobs required largely similar work but differed in terms of the status levels associated with the jobs, Homans asserts: "...the higher the person's rank, the closer his activities will come to the norm, or, even more simply, *noblesse oblige*" (1950, p.141). Since rank in Homans' study was conferred by group affiliation—by having a particular job—rather than by individual characteristics, the conjecture applies to the current context. While suggestive, correlation is not causation. Because the direction of causation is at issue, the laboratory is the appropriate place to test Homans' conjecture.[8] The current study differs from Homans' classic in many ways, but most significantly by experimentally inducing status differences in the laboratory.

3 Experimental Design

Experimental participants were recruited from among undergraduates and staff at the University of California, Berkeley. During the course of the experiment, each participant played from ten to fifteen rounds of a standard trust game as well as ten to fifteen rounds of a costless signaling game hereafter referred to as the "truth game." *To minimize learning over rounds, participants were never informed of the outcomes of any game in any round.* At the end of the experiment, one round of game play was randomly selected to determine participants' earnings from the experiment. The experiment was programmed and conducted with the software z-Tree (Fischbacher 2007). All

[8]This phenomenon could also rationalize a wide array of seemingly unrelated or contradictory results concerning the effects of status in the experimental economics literature. For example, the hypothesis is consistent with the results in Ball, Eckel, Grossman and Zame (2001) where prices in two-sided auctions favor high status experimental participants, if profit-maximization is the norm that governs that situation. It is also consistent with increased giving by high status players in public goods games (Kumru and Vesterlund, 2008).

sessions were conducted in the X-lab facilities at the University of California, Berkeley in 2006. Session sizes varied from 16 to 20 participants. No individual participated in more than one session. All together, eight sessions were conducted and 144 participants participated.

The experiment consisted of two phases: an initial identity-inducement phase, followed by a game-playing phase. The identity inducement phase varied across treatment (S-ID) and control (ID-only) sessions, while the game-paying phase was identical in all sessions. Each subject participated in exactly one session.

3.1 Identity Inducement Phase

3.1.1 ID-only sessions

Participants were randomly assigned one of two colors: purple or orange. Colors were assigned by publicly drawing poker chips out of a canvas bag so that assignments were transparently random. Participants could see—and it was also explained to them—that purple and orange poker chips were equally likely to be drawn. Each color can be thought of as an identity.

All "purple" participants were seated on one side of the room and all "orange" participants were seated on the other side of the room. Approximately five feet of empty space in the middle of the room separated the two groups. Each participant was separated from his or her neighbors by an opaque divider, essentially creating a private cubicle for each participant. After being seated, participants put on wristbands matching their assigned colors which they were instructed to wear for the duration of the experiment. All participants maintained the same color assignment for the entire experiment. In ID-only sessions, this constituted the entire identity-inducement phase: after being seated and putting on wristbands, participants proceeded directly to the game-playing phase.

3.1.2 S-ID sessions

In S-ID sessions, the identity-inducement phase involved a status-inducement component. Status inducement was achieved using a procedure drawing heavily on previous experimental studies of status in the sense that high and low status groups were created randomly and status differences were reinforced by treating the high status group better (*cf.* Ball and Eckel, 1998). Specifically, after dividing participants into color groups exactly as described above, one of the colors was

publicly randomly chosen to signify high status. Status was assigned by putting one purple poker chip and one orange poker chip into an opaque canvas bag, shaking the bag in full view of all participants, and 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. Participants were then seated in the same manner as in ID-only sessions—groups on opposite sides of the room with individuals in private cubicles—and given group-colored wristbands to wear for the duration of the experiment. Each participant's own color-group, and therefore status level, remained the same for the duration of the experiment.

To mimic the trappings of status, status-specific seating arrangements and tasks were assigned. High status participants were made more comfortable by being seated three per row, while low status participants were seated five per row. In terms of tasks, high status participants enjoyed refreshments while low status participants worked.[9] Low status participants' work was boring and tedious: they were given a list of names alphabetized by last name and asked to re-alphabetize it by first name, by hand. Each low status participant was provided with their own list of names, and the task was completed individually.[10] Importantly, although all low status participants were assigned the same task in order to induce a sense of shared fate, neither communication nor cooperation with own-group or other-group members was allowed. Similarly, while high status participants shared a common experience, any interaction between them was forbidden—they enjoyed their refreshments alone in their cubicle. After approximately 10 minutes, materials were collected and participants proceeded directly to the game-playing phase.

In total, three ID-only sessions, involving 60 participants, and five S-ID sessions, involving 84 participants, were conducted. Summary information is provided in Table 1.

[9]The refreshments were modest, consisting of a glass of water or lemonade as well as some snacks—crackers, cheese and grapes.

[10]By reinforcing the random status assignment, I follow previous experimental work on status in the laboratory. For example, in Ball, Eckel, Grossman and Zame (2001) a random status assignment was reinforced and made more salient by having low status participants clap for high status participants.

3.2 Game-Playing Phase

The game-playing phase was identical in both versions of the experiment. All participants played 10-15 rounds of the trust game followed by 10-15 rounds of the truth game (described below).[11] Before each round, all participants were randomly paired and game roles were randomly assigned so that co-players and game roles could potentially change every round. Pairings were not restricted by color group: a participant might be paired with an "orange" player one round and a "purple" player the next. Each participant was informed of their co-player's color group, but otherwise pairings were completely anonymous.[12] Participants were informed that at the end of each session one round of each game would be randomly selected to determine their earnings.

The trust game (Berg, Dickhaut and McCabe, 1995) is a two-player sequential moves game of perfect information. Here, the first mover (sender) is endowed with \$7 while the second-mover (receiver) has no endowment.[13] The sender sends some, all or none of his or her endowment to the receiver. Each dollar sent is tripled by the experimenter before being allocated to the receiver. The receiver then returns some, all or none of this tripled amount to the sender, ending the game. The unique subgame perfect equilibrium of the trust game with purely selfish preferences involves the receiver returning nothing and the sender sending nothing.

The truth game is a two-player costless-signaling game that can be thought of as modeling the interaction between a used car salesman and a prospective buyer. Nature decides which type of car a salesman has available—either a reliable car or a lemon, each being equally likely. After privately observing the type of car he or she has for sale, the seller can send one of two (costless) messages to the buyer: "reliable," or "lemon." The buyer observes the seller's message and can take one of two actions—buy the car or walk away.[14] The buyer prefers buying the car only if it is 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. Thus, the buyer's and the seller's preferences are aligned only when the car is reliable. Monetary payoffs in the truth game (Table 2) depend

[11] The number of rounds varied across sessions because of time constraints.

[12] Even though participants were seated by color group, there were several people (8-10) in each color-group in each session making the assumption of anonymity still plausible.

[13] This differs slightly from the trust game in Berg, Dickhaut and McCabe (1995), where both senders and receivers were endowed with \$10.

[14] The descriptive terms used here (reliable, lemon, buyer, seller) and later in the analysis of the truth game are for expositional purposes only. The state space and message space were actually {Heads, Tails}, each buyer's action set was {Left, Right} and players were called "senders" and "receivers." See the Appendix for experimental instructions.

solely on the quality of the car and the buyer's decision. The payoffs are constructed to make the buyer indifferent between buying and walking away whenever a seller's message does not change the buyer's (correct) prior beliefs about car quality. All payoffs are chosen to be of the same order of magnitude in order to ameliorate concerns about observed coordination being due to non-strategic factors—e.g., payoff dominance or pure salience effects—rather than to preferences, information and utility maximization (see, e.g., Schelling, 1960; or the discussion in Crawford, Gneezy and Rottenstreich, 2008). The full game tree is presented in Figure 1.

The key feature of the truth game is that with purely selfish players messages are completely uninformative in equilibrium; on the other hand, if sellers' ideals prescribe that they *should not* lie and sellers care about their ideals then messages can convey information. Furthermore, one way buyers can be more generous is to simply buy more frequently irrespective of the seller's message. This potentially allows for the separation of ideals-based changes in behavior from a more general tendency toward generosity. Experimental instructions for the truth game and the trust game appear in the Appendix.

In total there are 760 observations for each game in the data: 300 observations for each game stemming from ID-only sessions, and 460 observations for each game from S-ID sessions. Since these totals reflect multiple rounds of game-play, care was taken to minimize possible dynamic effects such as learning or hedging across rounds (described below). These steps appear to have been successful as, overall, participants' actions in early rounds of each game did not differ significantly from their actions in later rounds. Therefore, in the analysis that follows data are pooled over all rounds for each game.

3.3 Why multiple rounds of game-play?

In choosing how to implement game-play, I faced a difficult design choice: whether to use a direct-response format or to use the more common strategy method. While the direct response method has the benefit of making each decision feel more "real" and consequential, it has a major drawback in terms of data generated: counterfactual decisions are not observed. On the other hand, the strategy method allows the researcher to observe each participant's complete contingent strategy but reported decisions may be the result of only "cold," more deliberative, processes and suffer to a greater extent from hypothetical bias (see the discussion in Brandts and Charness, 2011).

Ultimately, I chose to use the direct-response method. To partially allay its severe data disadvantages, I implemented multiple rounds of game-play for each game. To minimize the impact of repeated-game effects, three steps were taken: i) before each round, pairings were randomly and anonymously assigned and, within each pairing, roles were randomly determined so that neither pairings nor roles persisted across rounds; ii) only one round of each game was chosen, at random, to determine each participant's earnings to lessen hedging motives and opportunities across rounds; and iii) participants were given no feedback on any outcome of any game played until the very end of the experiment—and even then were only informed of the outcome of their own games in the round randomly selected to determine their earnings. This last step is meant to rule out, as much as possible, learning about the population of participants across rounds.

4 Theoretical Framework and Hypotheses

Before turning to the results, it will be useful to have hypotheses in mind. Because there is no consensus in the existing literature about what effect introducing a status difference between otherwise trivially distinct groups will have, the current inquiry is largely exploratory. However, as a rough guide on what to expect I turn to (eminent sociologist) Homans' classic insight: "...the higher the person's rank, the closer his activities will come to the norm, or, even more simply, *noblesse oblige*" (1950, p.141).

To transform Homans' broad "high status/high standards" hypothesis into testable predictions, I construct a theoretical framework in the spirit of Akerlof and Kranton (2000). I consider only a two-person setting as the experiment involves two-player games. For $k \in \{i, j\}$, denote an individual's action by a_k and his or her money earnings by $x_k(a_i, a_j)$. Assume individual k belongs to an observable "identity category," C_k, and that each identity category—or, simply, identity—prescribes an "ideal" action, $a^I_{C_k}$, for each decision situation. In addition to one's own group, ideals may depend the co-player's action or the co-player's identity category: $a^I_{C_k} = f(a_{-k}, C_k, C_{-k})$. Finally, I assume utility is increasing in own money earnings but decreasing in the distance between one's chosen action (a_k) and one's ideal action ($a^I_{C_k}$). A simple specification which captures this tradeoff is given by:

$$U_k = x_k(a_k, a_{-k}) - \alpha_k(a_k - a^I_{C_k})^2 \tag{1}$$

In Equation 1, pecuniary utility is linear in own money earnings. The tradeoff between money utility and identity utility is captured by an individual-specific parameter $\alpha_k \geq 0$. As α_k tends to infinity, an individual's behavior conforms perfectly to his or her ideals. The case of $\alpha_k = 0$ represents standard purely pecuniary preferences. In principle, α_k could vary by identity as well. I do not focus on this possibility since it adds little explanatory power, but I point out in a later calibration exercise where relaxing this assumption might be warranted.

Ideals in this framework generally prescribe *good* behavior in line with, but not necessarily identical to, the social norm. The key difference between social norms and ideals is that of scope: social norms represent a broad consensus about what constitutes ethical behavior irrespective of the individuals involved; ideals, on the other hand, may vary by identity by construction. Ideals provide a way to think about variation in behavior due to ethical considerations at an intermediate level—neither completely idiosyncratic nor totally invariant across individuals within a society. The framework is general enough to nest many existing models of ethical behavior. For instance, inequality aversion (Bolton and Ockenfels, 2000; Fehr and Schmidt, 1999) can be accommodated by assuming all identity categories prescribe a common ideal, a^I, which does not vary with the co-player's identity and is equal to the action which implements equal final money earnings for oneself and for one's co-player. Heterogeneous behavior in this case would have to be explained by individual heterogeneity in α_k.

Broadly speaking, considering a decision situation with a known social norm, Homans' hypothesis suggests that a high status identity entails more demanding ideals than a low status identity. To test this broad hypothesis, I exploit *a priori* knowledge about plausible social norms in the different roles and games participants play in my experiment to generate role-specific hypotheses about how status may affect behavior. Toward this end, label the high (low) status category H (L).

Consider first the trust game. Denote by s the sender's (player i's) action and denote by $r(s)$ and $r^I(s)$, respectively, the actual and ideal amount the receiver (player j) returns conditional on s—suppressing for now category subscripts. Previous research suggests that the social norm

governing trust game receivers' behavior is reciprocity. A widely-used measure of reciprocity in this context is the slope of the receiver's return function: $\frac{\partial}{\partial s}r(s)$. The slope essentially captures how much more generosity is engendered from receivers by a marginal increase in the sender's niceness, which is a form of reciprocal behavior. To uncover the relationship between ideal and observed return amounts, I first plug receiver's money earnings, $(3s - r)$, into Equation 1, yielding overall utility: $U_j(s, r) = 3s - r - \alpha_j(r - r^I(s))^2$. Differentiating with respect to r to maximize utility followed by algebraic manipulation implies that the internal portion of receiver's optimal return function, $0 \leq r^*(s) \leq 3s$, satisfies: $r^*(s) = r^I(s) - \frac{1}{2\alpha_j}$. Thus, the slope of a receiver's observed return function—observed reciprocity—should provide an informative measure of the slope of the receiver's ideal return function—ideal reciprocity. On the other hand, changes in the marginal rate of substitution between pecuniary and ethical concerns, α_j, should increase or decrease return amounts uniformly across s, leaving the slope of $r(s)$ unchanged.

If Homans' hypothesis is true and high status entails ideals that are closer to the social norm, then I would expect high status receivers' ideals to exhibit more reciprocity than low status ideals $(\frac{\partial}{\partial s}r^I_H(s) > \frac{\partial}{\partial s}r^I_L(s))$. In terms of behavior, high status receivers' return functions should be steeper than those of their low status counterparts: $\frac{\partial}{\partial s}r_H(s) > \frac{\partial}{\partial s}r_L(s)$.

Hypothesis 1: *The sensitivity of return amounts to send amounts, $\frac{\partial}{\partial s}r(s)$, will be larger for high status receivers than for low status receivers.*

Alternatively, if the social norm governing receiver behavior is not reciprocity but unconditional generosity, or if the primary effect of status is on the individual-level tradeoff between money and ethics as captured by α_j, then I would predict that for any given s, high status receivers will return more than low status receivers.

Hypothesis 1': *For each s, high status receivers will return more conditional on s than low status receivers.*

While my hypotheses will primarily focus on understanding how ideals differ by status and thus consider only S-ID data, I do construct one set of hypotheses related to across-treatment variation in reciprocity. One way to think about reciprocity is as a tool to provide the proper incentives for others to behave in a way that benefits a group when intrinsic incentives are otherwise insufficient.

Since reciprocal behavior can be personally costly, individuals may focus their reciprocity on groups whose outcomes they care about or feel responsible for. Which groups people care about may run the gamut from one's immediate family only ("amoral familism") to more abstractly-defined groups such as those induced here. In line with this interpretation of reciprocity, I predict that ID-only receivers will behave more reciprocally when matched with an in-group sender than with an out-group sender—they may simply feel more responsible for the outcomes of their own group. Comparing ID-only receivers behavior to S-ID receivers, in line with the spirit of *noblesse oblige* I hypothesize that high status extends the scope of responsibility so that high status receivers will be as reciprocal in all of their interactions as ID-only receivers matched with in-group senders. Again interpreting the sensitivity of trust game receivers' actions to senders' action as a rough measure of reciprocity, this leads to the following two predictions:

Hypothesis 2A: *In ID-only sessions, trust game receivers' return proportions will be more sensitive to in-group senders' actions than to out-group senders' actions .*

Hypothesis 2B: *In S-ID sessions, high status trust game receivers' behavior will not vary by senders' group affiliations and will exhibit as much reciprocity as ID-only receivers matched with in-group senders.*

What about trust game senders? A common assertion is that the relevant social norm in this role is full trust—hence the name "the trust game"—and that send amount, s, serves as at least a rough proxy for trust. So, assume a sender's ideal prescribes a particular send amount, s^I. Repeating the exercise above, I write down an expression for a sender's money earnings and plug this into the general utility specification given by Equation 1 to get: $U_i = 7 - s + r^*(s) - \alpha_i(s - s^I)^2$. Maximizing utility by differentiating with respect to s followed by algebraic manipulation yields an expression for the sender's optimal internal action: $s^* = s^I + \frac{1}{2\alpha_i}[\frac{\partial}{\partial s}r(s) - 1]$. Translating this into predictions on behavior, we see that observed send amounts should be above, below or equal to senders' ideal send amounts depending on whether the slope of their co-players' return functions are greater than, less than or equal to one, respectively. However, in each of these three cases a higher s^I implies a higher s^*. Because senders are randomly matched from a common pool of receivers, there should be no systematic differences across senders in the receivers they face. Putting all of

this together implies that if Homans' conjecture rings true and high status senders' ideals come closer to the social norm of full trust ($s_H^I > s_L^I$), this should induce more trusting behavior among high status senders than among their low status counterparts ($s_H^* > s_L^*$).

Hypothesis 3: *High status senders will send more on average than low status senders.*

Turning from the trust game to the truth game, consider first truth game sellers. The seller chooses a probability, p, of telling the truth conditional on the state of the world. Previous research suggests that in situations where an individual can conceivably profit at another's expense by lying, always telling the truth—i.e., choosing $p = 1$—is the social norm. If high status ideals come closer to this social norm than low status ideals, they should prescribe a higher truth-telling propensity: $p_H^I > p_L^I$.

What are the implications for seller behavior? Buyers strictly prefer buying (walking away) when they believe the car is more (less) likely to be reliable than to be a lemon. Consider an equilibrium in which buyers buy with probability one conditional on the message "reliable" and with zero probability conditional on the message "lemon."[15] Assume a seller's ideal prescribes a probability for truth-telling conditional on knowing the car is a lemon, p_l^I. Suppose that the ideal probability for truth-telling when the car is reliable is $p_r^I > p_l^I$. In fact, assume for simplicity that $p_r^I = 1$.[16] The seller's strategy entails choosing two probabilities: the probability of telling the truth given that the seller knows the car is reliable, p_r, and the truth-telling probability conditional on knowing the car is a lemon, p_l. Given the buyer's strategy, the seller's utility is:

$$U_j(p_r, p_l) = \begin{cases} p_r[12 - \alpha_j(p_r - 1)^2] + (1 - p_r)[10 - \alpha_j(p_r - 1)^2], & \text{if car is reliable;} \\ p_l[10 - \alpha_j(p_l - p_l^I)^2] + (1 - p_l)[12 - \alpha_j(p_l - p_l^I)^2], & \text{if car is a lemon.} \end{cases} \tag{2}$$

In words, suppose the car is reliable and the seller chooses to tell the truth with probability p_r. Then with probability p_r the seller sends the message "reliable" and the buyer buys. The seller's money earnings are 12. The seller loses identity utility, however, if the chosen truth-telling

[15] Other equilibria are possible, including a perverse equilibrium in which the buyer believes the car is reliable whenever observing the message "lemon." I discuss this equilibrium in the Appendix.

[16] Relaxing this last assumption makes the algebra slightly messier, and implies some lying even when the seller knows the car is reliable but adds little intuition.

probability, p_r, does not equal the truth-telling probability prescribed by the seller's ideal, $p_r^I = 1$. In this case, the loss in identity utility is equal to $\alpha_j(p_r - 1)^2$. With probability $(1 - p_r)$ the seller lies and sends the message "lemon," causing the buyer to walk away. In this case, the seller earns 10 and again loses an amount of identity utility equal to $\alpha_j(p_r - 1)^2$. It is easy to see that a utility-maximizing seller sets $p_r^* = p_r^I = 1$, since decreasing p_r always lowers expected monetary earnings *and* identity utility.

Next, consider p_l^*. It is clear from the second equation above that when the car is a lemon the seller faces a tradeoff between money earnings and identity utility. Increasing truth-telling decreases money earnings in the equilibrium under consideration, but may increase identity utility. Differentiating the expression for utility when the car is a lemon with respect to p_l yields an optimal truth-telling probability that, when internal, satisfies:

$$p_l^* = p_l^I - \frac{1}{2\alpha_j} \tag{3}$$

Consequently, how likely the seller is to tell the truth about having a lemon for sale, or conversely, lying in a way that benefits oneself at the expense of others, depends on both the ideal truth-telling probability as well as how much one cares about ethical concerns in general. Increasing either p_l^I or α_j results in more personally costly truth-telling that benefits others. Finally, to verify this is an equilibrium it must be the case that $p_l^* > 0$, since this together with $p_r^* = 1$ and the buyer's uniform prior over car quality guarantees that the buyer's posterior belief that the car is reliable after observing a signal of "reliable" is larger than one-half. Equation 3 makes it clear that this is in general not true, but rather only when $p_l^I > \frac{1}{2\alpha_j}$—i.e., either when ideals prescribe a sufficiently high amount of truth-telling or when individuals' are concerned enough about ethical concerns in general. In particular, if high status entails a higher ideal truth-telling probability, p_l^I, the following hypothesis should obtain:

Hypothesis 4: *In the truth game, high status sellers will be more likely to tell the truth than low status sellers.*

What can our identity framework tell us about buyers' behavior in the truth game? Predictions so far have relied on knowing the social norm and then predicting that high status moves ideals

closer to this social norm. However, there is no unambiguous social norm in this role: on the one hand, it may be *good* to trust all messages; on the other hand, excessive credulity may be judged imprudent. Consequently, I expect high status to have little predictive power on buyers' *general* propensity to believe. At the same time, if Hypothesis 4 holds true and buyers anticipate this, one would expect buyers to be more likely to believe high status sellers' messages. This leads to my final hypothesis:

Hypothesis 5: *Buyers will be more likely to buy (walk away) following a message of "reliable" ("lemon") from a high status seller than following the same message from a low status seller.*

5 Results

5.1 The trust game

5.1.1 Summary statistics

Similar to the original trust game experiment (Berg, Dickhaut and McCabe, 1995), senders sent about half of the maximum possible amount with the average amount sent being roughly equal in the ID-only and S-ID sessions ($3.59 and $3.87, respectively). Although means were similar, it should be noted that a Kolmogorov-Smirnov test rejects the null hypothesis that the *distributions* of amounts sent were the same across versions ($p = 0.007$). Kernel density plots of amounts sent are presented in Figure 2, where vertical bars indicate averages. Overall, sender behavior appears more polarized—sending all or nothing both being more frequent—in S-ID sessions.

For their part, receivers returned approximately 80 percent of the amount they were sent (78 percent in ID-only; 77 percent in S-ID), representing a small loss on average for senders.[17] This figure is again comparable to Berg, Dickhaut and McCabe (1995) where receivers returned 89.5% of the money they were sent on average.[18]

[17] Since money sent was tripled, the proportion returned can take values from 0 to 3. Excluded are observations where nothing was sent where the return ratio $\frac{\$_{returned}}{\$_{sent}}$ is undefined.

[18] These similarities are striking given the the differences in initial endowments in the current design not present in the original trust game experiments, already suggesting that distributional preferences alone cannot explain participant behavior.

5.1.2 Receivers' behavior

Let us consider trust game receivers' decisions first. Receivers' pecuniary incentives are stark: they effectively decide how to share a fixed-size pie as in the dictator game. The dominant purely selfish strategy is to always return nothing.

It is widely believed that an important normative concern governing trust game receivers' behavior is reciprocity: rewarding kind behavior with kind behavior and unkind behavior with unkind behavior. While my first hypothesis above relates to the responsiveness of return *amounts*, when moving to the data this presents an obvious econometric concern about heteroskedasticity since the feasibility constraint on return amount obviously varies over amount sent, i.e., $0 \leq r(s) \leq 3s$. I account for this concern in two ways. The first strategy is indirect. I begin by examining return *ratios*, which also capture the sensitivity of receivers' behavior to senders' behavior but are not subject to varying feasibility constraints: $\forall s,\ 0 \leq \frac{r(s)}{s} \leq 3$. My second, more direct but less familiar, strategy is to use a semi-parametric estimation technique that is robust to the type of challenges posed by varying feasibility constraints. This strategy permits me to estimate receivers' return amount functions directly. It has the added benefit of allowing me to provide an estimate of receivers' ideal return amount strategies. The results from both analyses support Hypothesis 1.

Result 1: *Hypothesis 1 is supported. High status receivers are more reciprocal than low status receivers.*

I start by examining return ratio functions. Table 3 presents estimates of receivers' return ratios using data from the S-ID sessions. The estimates reveal that return ratios were significantly more sensitive to amounts sent for high status receivers than for low status receivers. Estimates in the last column of Table 3 suggest that, controlling for the status category of the sender, high status receivers were about three times as reciprocal as low status receivers.[19] Moreover, senders' status levels were irrelevant to both high and low status receivers: none of the coefficients on sender status or on an interaction with sender status are individually or jointly significant.[20]

[19]Specifically, the marginal impact of a dollar sent on the average return ratio was 0.043 for low status receivers, while it was $0.043 + 0.106 = 0.149$ for high status receivers—more than three times as large. Furthermore, in all but the least elaborate specification (column 1) the estimated marginal impact of one additional dollar sent on low status receivers' return ratios is not significantly different from zero, indicating a strikingly weak concern for reciprocity on the part of low status receivers.

[20]The following robustness checks were conducted (not reported, but available upon request): i) including individual

Next, I estimate receivers' return *amount* functions again restricting attention to S-ID sessions. To be econometrically rigorous, I exploit the fact noted above that a receiver's optimal internal return amount strategy can be written as: $r^*(s) = r^I(s) - \frac{1}{2\alpha_j}$. To simplify matters, assume that ideals are linear in s and that there is 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 α_j.[21] To further simplify matters, assume that all α_j's are drawn from a common distribution as would be expected if this parameter captures a stable, individual trait. Call this random variable α.

Given these assumptions, a straightforward way to estimate return ratios would be to use the expression for r^* above directly as a regression function with $\frac{1}{2\alpha_j}$ serving as an error term. Since r^* is possibly censored—whenever r^* falls below zero, I observe zero[22]—and it is unclear what distribution the error term should have, I use 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).[23] It is sufficient, for instance, to assume that $\frac{1}{2\alpha}$ is a well-defined random variable with a unique finite median. The CLAD estimate of receivers' optimal return function is presented in Table 4, column 1. As a robustness check, I also present results in column 2 of the same table using an alternative, more commonly used, estimator which also accounts for censoring (Tobit) but relies on stronger error-term assumptions. Senders' status again had no significant impact on receivers' return amounts. For ease of interpretation, I consequently report only estimates without controls for senders' status. Both estimates paint a similar picture: the positive and significant coefficient on the interaction between receiver status and send amount indicates that high status receivers are substantially more reciprocal than low status receivers.

Examining receivers' reciprocity by either measure, return amounts or ratios, ratios, leads to a

receiver random effects; ii) clustering standard errors at the individual level rather than the session level; iii) accounting for censoring using Tobit instead of OLS. In all of these alternate specifications: i) high status receivers are significantly more reciprocal than low status receivers; and ii) senders' group affiliation did not matter for behavior.

[21] It might seem more natural to explain variation in behavior with individual heterogeneity in the weights placed on identity utility, α_j, and ideals that do not vary by identity category. However, with this approach it is difficult to simultaneously explain why high status receivers are less generous when sent a low amount and more generous when sent a high amount—a result that obtains below.

[22] 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.

[23] 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 the data. Roughly speaking, there must be "enough" uncensored observations which is likely to be satisfied in the present case.

second result:

Result 1': *Hypothesis 1' is not supported. High status receivers are not uniformly more generous than low status receivers.*

Both Table 3 and Table 4 suggest that high status receivers could be either more or *less* generous than low status receivers. The negative and significant coefficient on the dummy for high status receivers in both tables indicates that for low send amounts high status receivers actually returned significantly less money than low status receivers. This latter feature is also apparent in Figure 3 (left panel) which plots raw return ratio data and simple linear fits for high status and low status S-ID receivers separately.

Before moving on, it should be noted that the weakness of the effect of *sender* status here is important because it makes the results difficult to explain with distributional equity concerns alone. That is, one reason high and low status receivers' behavior could differ is because the status manipulation simply introduced an unequal wealth distribution—the low status group "earned less" by not receiving refreshments. This explanation is made less plausible by the fact that all receivers apparently ignored whether they were returning money to a "poorer" participant.

5.1.3 Comparing receivers' behavior: ID-only vs. S-ID

Comparing reciprocity patterns across the experimental conditions yields insights into the working of status, leading to my next result:

Result 2: *Both hypotheses 2A and 2B find support in the data. ID-only receivers are more reciprocal when matched with an in-group sender than when matched with an out-group sender (2A). In S-ID sessions, high status receivers' behavior irrespective of the sender's group is remarkably similar to that of ID-only receivers matched with an in-group sender (2B).*

Reciprocity patterns induced by high status among S-ID receivers were qualitatively similar to reciprocity patterns in ID-only sessions—but with high status serving the role of in-group affiliation. In particular, estimating an individual random effects model of return ratios as a function of

amounts sent for ID-only receivers reveals that receivers' return ratios are (marginally) significantly more responsive to amounts sent for in-group senders than for out-group senders (Table 5).[24]

The similarities between the effects of high (low) status in the context of unequal-status groups (S-ID) and in-group (different-group) pairings in ID-only sessions are evident when comparing the left and right panels of Figure 3, which presents scatter plots of the raw data overlaid with simple linear fits for S-ID and ID-only receivers separately.[25] Qualitatively, the most striking similarities are: i) as already mentioned, steeper return ratio functions are associated with high status receivers (left panel) and same-color pairings (right panel); ii) ID-only receivers are less generous to in-group senders for low send amounts, mirroring the behavior of high status receivers in the S-ID data; iii) there is a break-even level of investment—approximately \$6—when receivers are either high status (S-ID) or when the sender and receiver share the same color (ID-Only); and, finally, iv) when receivers are either low status (S-ID) or from a different group than the sender (ID-only) there is no level of trust/investment yielding a positive "return on investment."

A bit more formally, Figure 4 presents plots of the raw data and fitted values for two sets of comparisons. The left panel compares high status S-ID receivers to same-group ID-only pairs; the right panel of Figure 4 compares low status S-ID receivers to different-group ID-only pairs. The shaded region in each panel represents a 95 percent confidence interval for fitted values using data from S-ID sessions. The confidence intervals use non-clustered standard errors, so they should be viewed as unnecessarily tight bounds. Still, even with these tight confidence intervals it is apparent that neither the pair of return ratio functions in the left panel, nor the pair in the right panel, differ significantly from one another.

5.1.4 An Estimate of Receivers' Ideals

The simplicity of the strategic situation faced by receivers in the trust game provides an opportunity to estimate participants' ideals. In this role how much to return is purely a choice between ideals and self-interest. I further restrict my attention to trust game receivers in the S-ID version of the experiment where I have the most, and the cleanest, data. Denoting senders' actions by s, receivers'

[24] Individual random effects are used to take into account the fact that we have repeated observations on individuals where in some observations the individual is part of an in-group pairing and in some other observations the same individual is involved in an out-group pairing.

[25] The regressions for the S-ID data roughly correspond to Table 3, column 1, as they do not control for sender status. The figure therefore represents a conservative estimate of how return ratio function slopes vary by status.

actions by r and money payoffs by $x(s,r)$, receiver $j's$ overall utility is given by the Equation below, where α_j is a parameter capturing how much agent j cares about identity relative to pecuniary incentives:

$$U_j(s,r) = x_j(s,r) - \alpha_j(r - r^I_{C_j}(s))^2 \tag{4}$$

In Equation 4, agent j cares about both her money earnings—$x_j(r,s)$—and the distance between her actual action, r, and her *ideal* action, $r^I_{C_j}(s)$. Agent $j's$ ideal, in turn, depends on s—to incorporate reciprocity—as well as her social category, C_j.[26]

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

$$U_j(s,r) = x_j(s,r) - \alpha_j(x_j(s,r) - x_i(s,r))^2 \tag{5}$$

Thus, receiver j 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 (5) and simplifying yields:

$$U_j(s,r) = 3s - r - \alpha_j(4s - 2r - 7)^2 \tag{6}$$

From Equation 6, simple algebraic manipulation allows one to re-cast the model of inequality aversion given by Equation 5 in an identity-utility form:

$$U_j(s,r) = x_j(s,r) - \tilde{\alpha}_j(r - r^I(s))^2 \tag{7}$$

In Equation 7, $r^I(s) = 2s - \frac{7}{2}$ and $\tilde{\alpha} = 4\alpha$.[27] Thus, inequality aversion can be thought of as

[26]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 modeled.

[27]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 into an $r - r^{Ideal}(s)$ format. This is also where the $\frac{7}{2}$ term in r^{Ideal} comes from.

an identity model which adds two assumptions: i) ideals are constant across social identities; and ii) ideals take a specific form. Because of this connection, I will use as my null hypothesis the idea that the data in the current experiment are well-explained by inequality aversion.

Proceeding with the estimation of receivers' ideals, after plugging the receiver's money earnings, $3s - r$, into the identity model of Equation 4 first-order conditions imply that a receiver's optimal (interior) money-return rule is given by:

$$r_j^*(s) = r_{C_j}^I(s) - \frac{1}{2\alpha_j} \tag{8}$$

Thus, estimating a receivers' average optimal return rule, $r^*(s)$, is equivalent—up to an unknown constant—to estimating receivers' ideals. To simplify matters, assume that ideals are linear in s and that there is precisely one ideal for each status level—$r_L^I(s)$ for low status receivers, and $r_H^I(s)$ for high status receivers. Call the CLAD estimates of receivers' optimal return rules among high (low) status participants from Table 4 $\widehat{r_H^*}$ ($\widehat{r_L^*}$). Estimating receivers' ideals is a matter of simply shuffling terms from one side of Equation 8 to the other. Specifically:

$$\widehat{r_H^I} = \widehat{r_H^*} + \frac{1}{2\overline{\alpha}} \tag{9}$$

$$\widehat{r_L^I} = \widehat{r_L^*} + \frac{1}{2\overline{\alpha}} \tag{10}$$

Plugging into Equations 9 and 10 the estimates of $\widehat{r_H^*}$ and $\widehat{r_L^*}$ from Table 4 yields the estimated ideals:

$$\widehat{r_H^I} = 2.37s - 6.88 + \frac{1}{2\overline{\alpha}}$$

$$\widehat{r_L^I} = 0.57s - 0.00 + \frac{1}{2\overline{\alpha}}$$

Notice that high status participants' ideals are close to the ideals of inequality-averse agents: $r^I(s) = 2s - \frac{7}{2}$. Furthermore, recall that high status participants behave like participants in same-group pairings *sans* status inequality. This suggests one reason participants may *appear* inequality averse in laboratory experiments, but not outside of the lab, is that merely bringing participants

into the lab creates a shared social identity.[28] It also suggests that one reason non-findings can occur in social identity experiments with weak identity-inducement procedures is a failure to eliminate this unintended shared social identity. Finally, the fact that low status participants' ideals do not resemble the ideals associated with inequality aversion suggests that I can reject the null hypothesis of inequality aversion: ideals are not constant across identities.

We can already see that high status receivers' ideals are more sensitive to senders' actions. To justify the interpretation that this implies more reciprocity with respect to conformance to ideals, for ease of exposition assume that senders' ideals do not vary with identity and, moreover, prescribe full trust: $s_H^I = s_L^I = 7$. Re-write receivers' ideals as in Equation 11 below where γ captures a base-line level of generosity, while β measures punishment or reward norm conformance:

$$r^I(s) = \beta(s - s^I) + \gamma \tag{11}$$

Receivers' estimated ideals can be re-written:

$$r_H^I(s) = 2.37(s - s^I) + (9.71 + \frac{1}{2\overline{\alpha}}) \tag{12}$$

$$r_L^I(s) = 0.57(s - s^I) + (3.99 + \frac{1}{2\overline{\alpha}}) \tag{13}$$

$$\tag{14}$$

In this form, we can see that 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: the average *ideal* return amount when senders exactly conform to their ideal, $(s - s^I) = 0$, is $\$[9.71 + \frac{1}{2\overline{\alpha}}] - \$[3.99 + \frac{1}{2\overline{\alpha}}] = \$5.72$ higher for high status receivers than for low status receivers.[29]

[28] For direct supporting evidence on this point in the context of a dictator game with third party punishment, see Butler, Conzo and Leroch (2013).

[29] The actual average difference in the data is \$1.53. This does not necessarily imply that my model organizes the data poorly, however. The observed discrepancy between the model prediction and the data may result from several factors: senders' ideals may vary by status, which I assume away for simplicity here; alternatively, α_j may have a group-contingent component, which I again have assumed away for simplicity. For example, the pattern is consistent with low status receivers generally placing less weight on identity ($\overline{\alpha_L} < \overline{\alpha_H}$).

5.1.5 Senders' behavior

It is widely believed that the guiding normative principle for senders in the trust game is trust—hence the name, "the trust game."[30] As detailed in the hypotheses section above, this implies the testable prediction that high status senders will exhibit more trust, sending more on average, than low status senders. This prediction is supported by the data.

Result 3: *Consistent with Hypothesis 3, high status senders send more to their co-players on average than low status senders.*

To show this, I assign to each individual his or her average amount sent—where this average is taken over all rounds in which an individual played the role of sender. This yields 84 observations in total for the S-ID sessions. To justify this procedure, notice that since there was little possibility for learning each sender faced essentially an identical choice each round.[31] Furthermore, since pairings were randomly re-determined each round, there should be no bias in the status of the receivers individuals were paired with, making this measure of trust comparable across individuals. Constructing one average measure of trust for each individual is therefore a transparent way of addressing concerns about artificially increasing the significance of statistical tests through repeated observations on the same individuals.

Using this measure of each individual's average trust level, high status senders were significantly more trusting than low status senders. In fact, on average high status senders sent *27 percent more* than low status senders.[32] Both high and low status receivers were the beneficiaries of high status senders' enhanced trust—so much so that high status S-ID senders displayed more trust in low status receivers than even their own (low status) group members did. On average, high status senders sent \$4.19 to low status receivers compared with just \$3.54 sent by low status senders.[33]

[30] There is an important subset of economists who argue that the trust game has nothing to do with trust. Addressing this meta-critique of the trust game is beyond the scope of the current inquiry, but direct evidence that the trust game *does* involve trust is provided in Butler, Giuliano and Guiso (2012).

[31] This is not true for receivers, of course, because each round a receiver faces a potentially different choice: how much to return conditional on being sent x dollars, $x = 0, \ldots, 7$. Therefore, the simpler route of constructing one average reciprocity measure for each individual receiver would be problematic.

[32] On average, high status senders sent \$4.26, while low status senders sent \$3.47. Using a one-tailed t-test the difference is significant at the 5 percent level ($p = 0.042$). Regressing individuals' average send amounts on a dummy for high status senders yields a marginally significant difference ($p = 0.064$). Both of these tests involve 84 observations—one for each S-ID participant. The OLS regression uses robust standard errors clustered by session.

[33] Again, these numbers are calculated using one (summary) observation per individual, for a total of 84 observations.

5.1.6 Comparing senders' behavior: ID-only vs. S-ID

As with receivers, high status S-ID senders resemble ID-only senders engaged in same-group pairings. Senders in the ID-only sessions sent more to in-group members than to out-group members, just as high status S-ID senders sent more to everybody than low status S-ID senders. And, again, the effect of in-group affiliation was weaker than the effect of status. ID-only senders sent 10.6% more, on average, to in-group members than to out-group members—a figure which is roughly one-third the size of the 27% trust increase exhibited by high status S-ID senders relative to their low status counterparts reported above.[34] The raw data on sender behavior is reported in Table 6.

To summarize, patterns in the trust game are consistent with the "high status/high standards" hypothesis. High status participants' behavior mirrored more closely the governing moral principle in both the sender and receiver roles of the trust game than did their low status counterparts. Additionally, high status participants' behavior in S-ID sessions qualitatively mirrored behavior associated with in-group pairings in an equal-but-different-groups context (ID-only). In terms of specific predictions, Hypotheses 1, 2A, 2B and 3 were supported by the data, while Hypothesis 1' was not.

Let us now consider whether these patterns carry over to a situation guided by different moral concerns: the truth game.

5.2 The truth game

Recall that the Truth game is a two-player costless signaling game that can be thought of as modeling the interaction between a seller with private information about the quality of a car, and a buyer who has no information besides the seller's costless message (for payoffs, see Table 2; for game tree, see Figure 1). Because interests are misaligned and the message space is the state space, seller honesty is defined and potentially helpful to the buyer but not possible in equilibrium with purely selfish players. On the other hand, if the seller's preferences include disutility from dishonesty, messages can be both informative and beneficial to buyers in equilibrium.

[34]The increased trust exhibited by high status S-ID senders in the raw data is slightly smaller, at 22%, but still considerably larger than in-group bias in ID-only sessions.

5.2.1 Summary statistics

Overall, at first glance sellers' messages appear quite similar in S-ID and ID-only sessions (Table 7). Upon closer inspection, however, there is about a five percentage point increase in honesty in S-ID relative to ID-only for each quality level: S-ID sellers were more likely to send the message "reliable" when the car was actually reliable (94.2 percent vs. 89.6 percent); and more likely to send the message "lemon" when the car was actually a lemon (34.3 percent vs. 29.5 percent). On the other side of the market, the data suggest that buyers conditioned their decisions on sellers' messages even though messages should be completely uninformative about car quality in any equilibrium with purely selfish players (Table 8). For example, buyers were more than twice as likely to "buy" after receiving the message "reliable" than after receiving the message "lemon." Taken together, these patterns already suggest both the importance of ideals for sellers and the anticipation of this importance by buyers.

5.2.2 Truth game sellers

Truth Game sellers' actions in the S-ID sessions are consistent with the high status/high standards hypothesis. For sellers, the norm is clear: a decent person *should* tell the truth. If high status sellers' ideals incorporate this norm to a greater extent, therefore, we would expect high status sellers to be more honest than low status sellers. In fact, this was true.

Result 4: *Hypothesis 4 is supported in the data. High status sellers were more honest than low status sellers.*

To show this, I estimate a logit model with honesty as the (binary) dependent variable (Table 9).[35] Controls include seller's status, seller's private information and an interaction between the two. A positive and significant coefficient on the dummy for high seller status shows that high status sellers were generally more honest. Overall, high status sellers were 9.3 percentage points more likely to tell the truth, which is about 15 percent of the sample mean—an economically and statistically significant difference.

[35]That is to say, this dummy variable takes the value of 1 when the seller's message matches his or her private information and 0 otherwise.

5.2.3 Comparing sellers' behavior: ID-only vs. S-ID

To compare seller honesty across treatment and control, Table 10 presents logit models of honesty for both S-ID and ID-only. Here, controls include a dummy variable for whether the seller and buyer shared the same color group, the private information of the seller and an interaction term. The estimates reveal a pattern similar to that found in the trust game data, above: a classical in-group bias when groups are equal-but-different (column 1) which is eliminated by the introduction of a salient status difference between groups (column 2), and replaced by better-behaved high status sellers. In ID-only sessions, in-group sellers are significantly more likely to tell the truth about bad private information—i.e., when the seller knows "lemon." The increase in honesty propensity is about 8 percentage points when sellers have bad private information in the ID-only sessions which constitutes approximately 30 percent of the sample mean in this situation, while there are no significant effects of being in an in-group pairing in S-ID sessions.

Table 10, column 1, also suggests ID-only sellers are marginally *less* likely to tell the truth about *good* private information when matched with an in-group seller than when matched with an out-group seller, as indicated by the (marginal) significance of the coefficient of the "Same-Group" indicator. At first glance this may seem puzzling. However, as discussed in the Appendix, this can be plausibly attributed to a perverse equilibrium with pooling on the message "lemon."[36] Excluding observations where sellers lie about good news, and hence may be involved in this perverse equilibrium, the similarity between ID-only sellers involved in in-group pairings and S-ID high status sellers is enhanced: i) high status sellers are 10.7 percentage points more likely to tell the truth about bad news than low status sellers; ii) ID-only sellers involved in in-group pairings are 12.1 percentage points more likely to be honest about bad news than those involved in different-group pairings.

5.2.4 Truth game buyers

There is no clear social norm for buyers. Consequently, the high status/high standards hypothesis makes no clear prediction about how buyers' own status levels affect their behavior. Reassuringly,

[36]This equilibrium arises when buyers always believe the *opposite* of the message sent. Given these "incredulous" buyers, sellers pool on sending the message "lemon" irrespective of their private information, so that sellers lie about good information and tell the truth about bad information.

a buyer's own status has no significant impact on his or her buying decisions (Table 11, top panel). Continuing the theme of high status mirroring the effect of in-group pairings *sans* status differences, it is also the case that in ID-only sessions being matched with an in-group seller does not affect the buyer's general propensity to buy (Table 11, bottom panel).

However, since participants may anticipate the impact of identity on sellers' behavior and consequently condition their beliefs and ultimately their buying decisions on the seller's identity, identity may still have an impact on equilibrium buying behavior. In particular, since high status sellers were generally significantly more honest than low status sellers one might expect buyers to more readily believe high status sellers' messages in S-ID sessions. This prediction turns out to be true.

Result 5: *Hypothesis 5 finds support in the data. Buyers' behavior is consistent with being more likely to believe messages sent by high status sellers.*

To show this, I construct an indicator variable—*believe*—to take the value one whenever the buyer takes an action consistent with believing the seller's message. Specifically, *believe* equals one if either: i) following a message of "lemon" the buyer chooses "walk"; or ii) following a message of "reliable" the buyer chooses "buy." Otherwise *believe* takes the value zero. Restricting attention to S-ID sessions, raw proportions for *believe* by buyer and seller status are presented in Table 12. High status sellers' messages were believed in over 80% of interactions regardless of the status level of the buyer, while low status sellers' messages were believed in only about 72% of interactions. Controlling for seller's status, there was little variation in seller credibility/buyer credulity over the buyer's status level. To show the statistical significance of this credibility gap, I restrict attention to S-ID sessions and estimate a logit model with *believe* as the dependent variable, a dummy for high seller status as the (lone) independent variable and with robust standard errors clustered by session (not reported but available upon request). The estimated coefficient on the high seller status dummy is positive and highly significant ($p = 0.000$), indicating that buyers were significantly more likely to believe high status sellers.

The lack of an in-group seller effect on credulity in ID-only sessions is evident in the raw data: ID-only buyers believe in-group sellers 71.3 percent of the time and out-group sellers slightly *more* frequently (75.3 percent of the time). What accounts for the lack of an in-group effect on buyer

behavior in ID-only sessions? While I have no definitive answer, it is worth noting that buyers' behavior there was less straightforward. Sellers matched with in-group buyers were not uniformly more honest, potentially because of an increased likelihood of playing a perverse equilibrium pooling on the message "lemon."

More generally, the fundamental question for buyers is not how likely they are to believe but, rather, how frequently they are able to make the "correct" choices: buying reliable cars and walking away from lemons. In the absence of messages—or equivalently, in any equilibrium with messages and purely selfish sellers—we would expect correct choices only 50 percent of the time. If, however, messages are informative *and* buyers correctly perceive and act upon the information content of the messages then we would expect more than 50 percent of choices to be correct. This leads to the final result of the paper.

Result 6: *Status differences benefit buyers.*

In S-ID sessions, buyers were 22 percent more likely to make the correct decision which is statistically and economically significant. Statistical significance is supported by various tests. Among these are the following. First of all, I estimate a logit model with "correct decision" as the binary dependent variable and a dummy for S-ID sessions as the independent variable. With robust standard errors clustered by session the coefficient on the dummy for S-ID sessions is significantly larger than zero ($p = 0.002$). Secondly, for a more conservative figure I compute, for each session, the average proportion of correct choices. This results in eight observations since there are eight sessions. Using these eight observations I perform a non-parametric one-tailed permutation test. Even according to this test the proportion of correct choices is higher in S-ID sessions at a close-to-standard level of significance ($p = 0.071$). Moreover, S-ID buyers were better off than ID-only buyers irrespective of the buyer's status level. Low status buyers chose correctly in 60.6 percent of their interactions, while high status buyers made the correct choice 59.4 percent of the time. In contrast, in the ID-only sessions where group affiliations did not change buyers' behavior buyers made the correct choice only 49 percent of the time—just as if there were no messages at all.

To compare the effects of status on outcomes across games, I construct a general measure of market efficiency. For each two-player interaction, I compute how much larger than its theoretical minimum the pair's realized total earnings were. I then normalize by the theoretical maximum

this difference could achieve: $efficiency = \frac{[\text{sum of earnings}]-[\text{min sum of earnings}]}{[\text{max sum of earnings}]-[\text{min sum of earnings}]}$.[37] This results in a measure of efficiency which can take values between 0 and 1 for each of the two games considered.[38] To make the comparison across games as meaningful as possible, in the truth game I restrict attention to those situations where players' decisions could affect efficiency: interactions in which the car for sale was actually reliable. When the car was a lemon all outcomes featured the same level of efficiency: total earnings are always $10 + 12 = 22$.

With this caveat in mind, introducing status increased efficiency in both games. Considering the 358 truth game interactions featuring a reliable car, S-ID sessions featured an average efficiency level of 0.77 while ID-only sessions' efficiency was significantly lower at 0.67 ($p = 0.027$, one-tailed permutation test). Introducing status also increased efficiency in the trust game, although the increase was milder. S-ID trust game efficiency was 0.55, on average while ID-only sessions achieved an efficiency level of 0.51 ($p = 0.087$; one-tailed permutation test). Restricting attention to S-ID sessions and considering how efficiency was affected by the seller's status level reveals that interactions with high status truth game sellers achieved higher efficiency than those with low status sellers (0.79 vs. 0.76) as did trust game interactions with high status senders compared to low status senders (0.61 vs. 0.50). The latter difference is statistically significant ($p < 0.01$; one-tailed permutation test). Considering ID-only sessions separately, in-group pairings were more efficient on average. In the truth game, interactions involving a reliable car achieved an efficiency measure of 0.69 when the seller and buyer shared a group affiliation, but only 0.65 when they did not. In the trust game, same-group pairings were associated with an efficiency level of 0.54 vs. 0.49 among different-group pairings. Neither of these latter two efficiency differences were statistically significant, however. All together, the results in terms of efficiency are qualitatively similar to the patterns in the raw data: in-group pairings as well as pairings involving a high status first-mover were more efficient and, moreover, introducing status differences among otherwise trivially distinct groups enhanced market efficiency.

To summarize the results so far, the data suggest that high status causes closer adherence

[37] I thank an anonymous referee for suggesting this comparison, generally, and this efficiency measure specifically.

[38] As an example, consider an interaction in which the trust game sender sends 3. The total earnings of both players would be $7 + 2 * 3 = 13$ since each dollar sent creates two extra dollars in surplus. Maximum total earnings occurs when the sender sends as much as possible, i.e., 7, yielding total earnings of $7 + 2 * 7 = 21$. The minimum possible total earnings (7) occurs when the sender sends 0. Thus, the efficiency of an interaction where a sender sends 3 is $\frac{13-7}{21-7} \approx 0.43$.

to relevant social norms, producing more trust and reciprocity among trust game senders and receivers, respectively, and more honesty among high status sellers in the truth game. There is no obvious social norm for truth game buyers, and, consequently, behavior does not vary by own status. However, truth game buyers correctly anticipated that high status would produce more honesty in the truth game which allowed buyers to benefit from the introduction of status differences. Finally, in both the truth game and the trust game there is a close connection between the effect of high status and the effect of in-group affiliation. Next, let us consider whether two obvious alternative explanations explain the data well.

6 What About Wealth Effects and Mood Effects?

There are two obvious alternative explanations for the patterns in the data: wealth effects and mood effects. Let us first consider wealth effects.

Could it be that the status manipulation in the S-ID version of the experiment simply made the "high status" participants feel wealthier? The precise effect of extra wealth would depend on the model of preferences one has in mind, but heuristically one might expect the "wealthier" participants to be simply more generous. The increased generosity of high status participants when sent a large initial amount is consistent with such a wealth effect. However, the relative *decrease* in generosity observed when the "wealthier" participants are sent low amounts seems contrary in spirit to the idea that high status participants simply feel wealthier.[39] Furthermore, it would be difficult to explain the close parallels in ID-only and S-ID receivers' behavior with wealth effects alone since there was no systematic "wealth" difference created in the ID-only version of the experiment. Finally, in the truth game, one obvious outlet for increased generosity would have been simply buying more irrespective of sellers' messages. In particular, one might expect high status (or, wealthier) buyers to have displayed more generosity towards low status sellers by buying more frequently from them. Neither of these patterns is in the data, however: high status buyers bought only slightly more frequently than low status buyers, overall (67.8 percent of the time vs.

[39]Pushing this intuition a bit further, if the proposed wealth effect resulted in an approximately uniform increase in generosity, we should see "wealthier participants" return a fixed amount more to their co-player for each amount initially sent 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 participants. This was exactly the *opposite* of the observed patterns.

65.1 percent, respectively); and high status buyers were *less* likely to buy from low status sellers than from high status sellers (66.2 percent of the time vs. 69.9 percent, respectively). All together, these patterns suggest wealth effects are unlikely to be the primary explanation for my data.

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 of the participants—the high status participants—in a better mood. At the same time, a growing body of experimental research in economics suggests that mood may have a significant impact on other-regarding behavior. Other extraneous factors plausibly operating through mood, such as whether an individual has just eaten lunch, may also effect one's general tendency to take others' interests into account (e.g., Danziger, Levav and Avnaim-Pesso, 2011). Without dismissing the potential effect of mood completely, I do have a few reasons for believing mood effects are not the primary drivers of my results. Firstly, I will reiterate that social identity had a significant impact even in the ID-only version of the experiment where it is less clear that moods varied. Secondly, there are two closely related experiments which manipulate mood that provide compelling evidence against mood as an explanation of the current results. In Kirchsteiger, Rigotti and Rustichini (2006), the authors induced two different moods—good or bad—by having participants watch either a funny or a sad movie, respectively. Participants then played a standard gift-exchange game.[40] The result of this manipulation was that participants in a good mood were *less* reciprocal and *more* generous than players in a bad mood. That is to say, players in a *bad* mood mirror high status participants in the present experiment; and players in a *good* mood mirror low status participants.[41] Capra (2004) produces similar findings using a trust game: inducing a bad mood leads to more trust and more reciprocity relative to inducing a good mood. Therefore, since the most likely mood effect of status in the current experiment was to induce a bad mood among low status participants, mood effects probably worked *against* the patterns in the data rather than providing an alternative explanation for them.

[40] 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.

[41] They also mirror, by the way, the hypothetically "wealthier" participants outlined above.

7 Conclusion

A necessary pre-condition for social identity to enrich economic models and yield a "...theory of decision making where social context matters..." (Akerlof and Kranton, 2010) is an understanding of how context interacts with identity to produce behavior. Because many economically-relevant contexts involve both equal and unequal social groups, an important first step in this direction is to examine how the introduction of group-level status inequality might *modify* the effects of social identity.

Consistent with previous research, in this paper I document in-group bias in the context of equal-but-different social identities. In contrast to previous research, however, the current inquiry provides novel evidence that in-group bias is a product of the equal-but-different social groups context. The data suggest that introducing status inequality between groups may weaken the impact of in-group/out-group distinctions and pave the way for other types of group-contingent preferences to color behavior. Specifically, in the context of *unequal* social groups members of high status groups may behave more in accordance with relevant social norms regardless of with whom they interact. This may operate through direct internalization of moral codes suggesting high status groups should act better, or indirectly through guilt aversion coupled with the belief that others believe high status groups should behave better. Taken together, my results suggest that identity affects behavior through social norm compliance and that *how* identity affects norm compliance depends on the nature of the relationship between the groups involved.

The current experiment raises important questions which are left for future research. Firstly, the social identities induced in this experiment were 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 dictated the associational patterns: with whom participants interacted was (randomly) assigned. However, one might suspect that, for instance, expected reciprocity patterns might have a significant impact on with whom participants would *choose* to interact. Such associational biases may have significant economic consequences, and are worthy of investigation.

References

[1] Akerlof, G. (1980), "A theory of social custom, of which unemployment may be one consequence," *Quarterly Journal of Economics*, 94: 749-775.

[2] Ball, Sheryl and Catherine Eckel (1998), "The economic value of status," *Journal of Socio-Economics*, 27: 495–514.

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

[4] Banfield, Edward C. (1958). *The Moral Basis of a Backward Society.* The Free Press, Glencoe, Il.

[5] Battigalli, Pierpaolo and Martin Dufwenberg (2007), "Guilt in games," *American Economic Review*, 97: 170–176.

[6] Ben-Ner, Avner, Brian P. McCall, Massoud Stephane, and Hua Wang (2009), "Identity and in-group/out-group differentiation in work and giving behaviors: Experimental evidence," *Journal of Economic Behavior and Organization*, 72: 153–170.

[7] Benjamin, Daniel, James Choi, and A. Joshua Strickland (2010), "Social identity and preferences," *American Economic Review*, 100: 1913–1928.

[8] Bernhard, Helen, Urs Fischbacher, and Ernst Fehr (2006),"Parochial altruism in humans," *Nature*, 442: 912–915.

[9] Bolton, Gary E. and Axel Ockenfels (2000). "ERC: A Theory of Equity, Reciprocity, and Competition," *The American Economic Review* 90: 166-193.

[10] Brandts, Jordi and Gary Charness (2011), "The strategy versus the direct-response method: a first survey of experimental comparisons," *Experimental Economics*, 14: 375–398.

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

[12] Butler, Jeffrey V., Paola Giuliano, and Luigi Guiso (2012), "Trust and cheating," *NBER Working Paper no. 18509*.

[13] Butler, Jeffrey V., Pierluigi Conzo and Martin Leroch (2013), "Social Identity and Punishment," *EIEF Working Paper no. 1316*.

[14] Capra, Monica (2004), "Mood-driven behavior in strategic interactions," *American Economic Review*, 94: 367–371.

[15] Charness, Gary and Martin Dufwenberg (2006), "Promises and partnership," *Econometrica*, 74: 1579–1601.

[16] Charness, Gary, Luca Rigotti, and Aldo Rustichini (2007), "Individual behavior and group membership," *American Economic Review*, 97: 1340–1352.

[17] Chen, Roy and Yan Chen (2011), "The potential of social identity for equilibrium selection," *American Economic Review*, 101: 2562–2589.

[18] Chen, Yan and Xin Li (2009), "Group identity and social preferences," *American Economic Review*, 99: 431–457.

[19] Chen, Yan, Tracy Xiao Liu, Yan Chen, Sherry Xin Li and Margaret Shih (2010), "Which hat to wear? impact of natural identities on coordination and cooperation," *University of Michigan Working Paper*.

[20] Crawford, Vincent P., Uri Gneezy and Yuval Rottenstreich (2008), "The Power of Focal Points Is Limited: Even Minute Payoff Asymmetry May Yield Large Coordination Failures." *American Economic Review*, 98: 1443–58.

[21] Danziger, Shai, Jonathan Levav and Liora Avnaim-Pesso (2011), "Extraneous factors in judicial decisions." *Proceedings of the National Academy of Sciences*, 108: 6889–6892

[22] Eckel, Catherine and Phillip Grossman (2005), "Managing diversity by creating team identity," *Journal of Economic Behavior and Organization*, 58:371–392.

[23] Fehr, Ernst and Klaus Schmidt (1999). "A Theory of Fairness, Competition, and Cooperation," *The Quarterly Journal of Economics*, 114(3): 817–868.

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

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

[26] Gneezy, Uri (2005), "Deception: The role of consequences," *American Economic Review*, 95: 384–394.

[27] Guala, Francesco, Luigi Mittone, and Matteo Ploner (2013), "Group membership, team preferences, and expectations," *Journal of Economic Behavior and Organization*, 86: 183–190.

[28] Güth, Werner, M. Vittoria Levati, and Matteo Ploner (2008), "Social identity and trust: an experimental investigation," *Journal of Socio-Economics*, 37: 1293–1308.

[29] Hargreaves Heap, Shaun P. and Daniel J. Zizzo (2009), "The Value of Groups," *The American Economic Review*, 99: 295–323.

[30] Harris, D., B. Herrmann, and A. Kontoleon (2010), "What is the nature and social norm within the context of in-group favouritism?," *Cambridge Working Papers in Economics no. (1062).*

[31] Hertel, G. and N. L. Kerr (2001), "Priming in-group favoritism: The impact of normative scripts in the minimal group paradigm," *Journal of Experimental Social Psychology*, 37: 316–324.

[32] Hoff Karla and Priyanka Pandey (2006), "Discrimination, social identity, and durable inequalities," *American Economic Review*, 96: 206–211.

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

[34] Jetten, Jolanda, Russell Spears and Antony S. R. Manstead (1997), "Strength of identification and intergroup differentiation: the influence of group norms," *European Journal of Social Psychology*, 27: 603–609.

[35] Jost, John T., Mahrzarin R. Banaji and Brian A. Nosek (2004), "A decade of system justification theory: accumulated evidence of conscious and unconscious bolstering of the status quo," *Political Psychology*, 25: 881–918.

[36] Kimbrough, Erik O. and Alexander Vostroknutov, (2013), "Norms Make Preferences Social," *Discussion Papers dp13-01*, Department of Economics, Simon Fraser University.

[37] Kirchsteiger, Georg, Luca Rigotti and Aldo Rustichini (2006), "Your morals might be your moods," *Journal of Economic Behavior and Organization*, 59:155–172.

[38] Krupka, Erin L. and Roberto A. Weber (2013), "Identifying social norms using coordination games: Why does dictator game sharing vary?," *Journal of the European Economic Association*, 11: 495–524.

[39] Kumru, C. and Lise Vesterlund (2008), "The effect of status on voluntary contribution," *UNSW Australian School of Business Research Paper No. 2008 ECON 02.*

[40] McLeish, Kendra N. and Robert J. Oxoby (2007), "Identity, cooperation, and punishment," *IZA Discussion Paper no. 2572.*

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

[42] Ostrom, Elinor (2000), "Collective Action and the Evolution of Social Norms," *The Journal of Economic Perspectives*, 14: 137-158.

[43] Schelling, Thomas (1960). *The Strategy of Conflict*, Cambridge, MA: Harvard University Press.

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

[45] Tsutsui, Kei and Daniel J. Zizzo (*forthcoming*), "Group Status, Minorities and Trust," *Experimental Economics.*

[46] Willer, Robb (2009), "Groups reward individual sacrifice: The status solution to the collective action problem," *American Sociological Review*, 74: 23–43.

Tables and Figures

Table 1: Summary of Experimental Sessions

	Sessions	Participants	Observations	Groups	Status Difference?	Learning?
ID-Only	3	60	300	Purple, Orange	No	minimal
S-ID	5	84	460	Purple, Orange	Yes	minimal

Table 2: Truth Game Payoffs

	Buyer Action	
	Buy	Walk Away
Actual Quality		
Reliable	$(12, 12)$	$(10, 10)$
Lemon	$(12, 10)$	$(10, 12)$

(a,b) = ($ Seller, $ Buyer)

Table 3: Trust game receivers, S-ID

Dependent Variable = Return Ratio = $\frac{\$Returned}{\$Sent}$			
	(1)	(2)	(3)
\$ Sent	0.056**	0.052	0.043
	(0.011)	(0.025)	(0.034)
High Status Receiver	−0.355**	−0.357**	−0.447*
	(0.101)	(0.097)	(0.175)
High Stat Rec ×\$*Sent*	0.086***	0.086***	0.106***
	(0.017)	(0.016)	(0.030)
High Stat Sender		-0.043	-0.146
		(0.306)	(0.230)
High Stat Sender ×\$*Sent*		0.008	0.030
		(0.056)	(0.075)
H.S. Sender × H.S. Rec'r			0.229
			(0.250)
H.S. Sender × H.S. Rec'r ×\$*Sent*			-0.047
			(0.053)
Constant	0.451*	0.469	0.513
	(0.173)	(0.254)	(0.297)
N	355	355	355
R^2	0.107	0.108	0.109

1. Dependent variable—return proportion—can take values from 0 to 3.
2. Robust standard errors, clustered by session, in parentheses.
3. Estimates include only observation where \$*Sent* > 0.
4. * indicates $p <= 0.10$; ** indicates $p <= 0.05$;*** indicates $p <= 0.01$

Table 4: Optimal Return functions, S-ID

Dependent Variable = Dollars Returned		
	CLAD	Tobit
$ Sent by sender	0.57^{***}	1.69^{***}
	(0.359)	(0.124)
High Status Receiver	-6.88^{***}	-1.69^{**}
	(4.665)	(2.805)
High Stat Rec'r $\times \$Sent$	1.80^{***}	0.50^{**}
	(0.880)	(0.202)
Constant	-0.00	-6.12^{***}
	(0.668)	(0.752)
$N_{initial}$	460	460
N_{final}	365	n/a
Censored Obs.	n/a	215
Pseudo R^2	0.301	0.146

1. Standard errors, clustered by session, in parentheses.
2. $*$ = significant at 10 percent; $**$ = significant at 5 percent; $***$ = significant at 1 percent.
3. For CLAD, significance tests use bias-corrected bootstrapped standard errors (not reported).
3. For Tobit, significance tests use robust standard errors.
4. Both CLAD and Tobit account for censoring when \$ Sent ≤ 0.

Table 5: Trust game receivers, ID-only

Dependent Variable = Return Ratio = $\frac{\$Returned}{\$Sent}$						
	\$ Sent	Same color sender	Same color sender × \$ Sent	Constant	Obs	R^2
Coeff	0.091^{***}	-0.040	0.022^{*}	0.327^{***}	249	0.077
StdErr	(0.018)	(0.035)	(0.013)	(0.050)		

1. Dependent variable—return ratio—can take values from 0 to 3.
2. Model incorporates individual-level random effects.
3. Robust standard errors, clustered by session, in parentheses.
4. Estimates include only observation where $\$Sent > 0$.
5. * indicates $p <= 0.10$; ** indicates $p <= 0.05$;*** indicates $p <= 0.01$

Table 6: Trust game senders' actions

Dollars Sent						
	ID-only			S-ID		
	overall	to in-group	to out-group	overall	to in-group	to out-group
Overall	3.59	3.77	3.41	3.87	4.04	3.72
	(0.271)	(0.444)	(0.134)	(0.348)	(0.388)	(0.342)
High Stat Sender				4.26	4.56	3.98
				(0.450)	(0.480)	(0.442)
Low Stat Sender				3.47	3.50	3.43
				(0.412)	(0.561)	(0.406)

1. Dollars sent can take values between 0 and 7.
2. Robust standard errors, clustered by session, in parentheses.

Table 7: Truth Game Sellers' Messages, overview

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 8: Truth Game Buyers, 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 9: Sellers' Honesty by Seller's Status, Logit Model.

Dependent Variable = Seller's Message Matched Private Information						
	High Status Seller	Lemon	Lemon×High Status Seller	Constant	N	Pseudo-R^2
Coeff	2.61**	−2.83***	-2.46	2.11	460	0.353
Std.Error	(1.187)	(0.439)	(1.680)	(0.295)		

1. Binary dependent variable takes the value 1 if the seller's message matched his/her private information and 0 otherwise.
2. Robust standard errors, clustered by session.
3. "Lemon" indicates seller knew car was unreliable.
4. * indicates $p <= 0.10$; ** indicates $p <= 0.05$;*** indicates $p <= 0.01$

Table 10: Sellers' Honesty by Nature of Pairing and Version, Logit Models.

Dependent Variable = Seller's Message Matched Private Information		
	ID-Only	S-ID
Same-Group	-0.79*	1.04
	(0.463)	(0.825)
Lemon	-3.66^{***}	-3.00^{***}
	(0.161)	(0.788)
Lemon×Same-Group	1.18^{***}	-1.22
	(0.354)	(1.035)
Constant	2.58^{***}	2.43^{***}
	(0.524)	(0.270)
N	300	460
Pseudo-R^2	0.299	0.338

1. Binary dependent variable takes the value 1 if the seller's message matched his/her private information and 0 otherwise.
2. * indicates $p <= 0.10$; ** indicates $p <= 0.05$;*** indicates $p <= 0.01$
3. Robust standard errors, clustered by session.
4. Same-group is an indicator for same-group pairings.
5. "Lemon" indicates seller knew car was unreliable.

Table 11: Buyers' decisions, Logit Model.

Dependent Variable = Buyer chose "buy"						
S-ID	High Status Buyer	Lemon	Lemon×High Status Buyer	Constant	N	Pseudo-R^2
Coeff	0.20	-2.63^{***}	-0.52	1.11^{*}	460	0.150
Std.Error	(0.725)	(0.671)	(0.788)	(0.577)		
ID-only	Same-color pair	Lemon	Lemon×Same group pair	Constant	N	Pseudo-R^2
Coeff	−0.10	-2.10^{***}	0.51	1.15^{***}	300	0.098
Std.Error	(0.280)	(0.346)	(0.326)	(0.148)		

1. Dependent variable takes the value of 1 if the buyer chose "Buy" and 0 otherwise.
2. Robust standard errors, clustered by session, in parentheses.
3. "Lemon" indicates buyer received message "lemon.".
4. * indicates $p <= 0.10$; ** indicates $p <= 0.05$;*** indicates $p <= 0.01$

Table 12: Truth Game Buyers' Credulity (S-ID)

Buyer's Propensity to Believe Seller's Message			
		S-ID	
		Buyer Status	
Seller Status	Overall	High	Low
High	0.814	0.825	0.805
Low	0.720	0.721	0.718

Figure 1: Truth Game, game tree

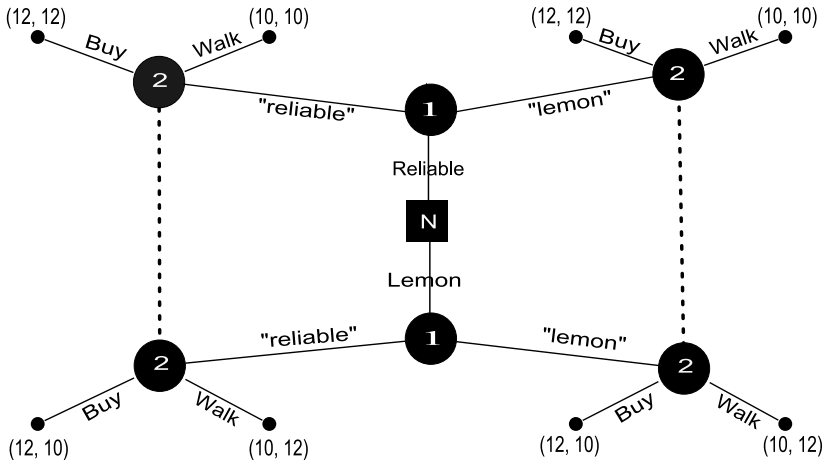

Figure 2: Amounts Sent, S-ID vs. ID-only

Figure 3: Trust game receivers, S-ID vs. ID-only

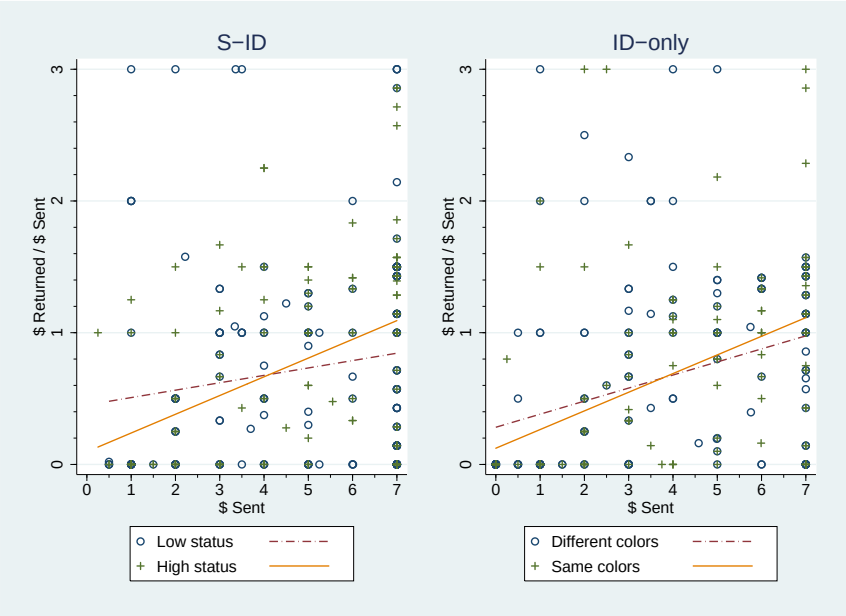

Figure 4: Return ratio functions comparisons